UNIVERSITY OF ALBERTA

Depth Estimation Error in Stereo Imaging

by

Hossein Sahabi ©

A Dissertation

Submitted to the Faculty of Graduate Studies

in patrial fulfilment of the requirements for the

degree of Doctor of Philosophy

Department of Electrical and Computer Engineering

Edmonton, Alberta

Fall, 2000

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59665-6

Canada

University of Alberta

Library Release Form

Name of Author: *Hossein Sahabi*

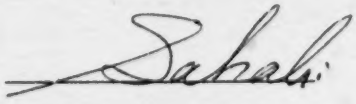
Title of Thesis: *Depth Estimation Error in Stereo Imaging*

Degree: *Doctor of Philosophy*

Year this Degree Granted: *2000*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Hossein Sahabi
Apt. B104, 2000 Walnut Ave,
Fremont, California, 94538
U.S.A.

Date: *August 18, 2000*

Abstract

In stereo imaging the depth of a 3D point is estimated based on the position of its projections on the left and right images. The imaging sensors of cameras, such as charge coupled devices (CCD), consist of discrete picture elements (pixels). The discretization of images generates uncertainty in estimation of the depth for each 3D point in the scene. In this thesis, the variation of depth estimation error in a typical stereo imaging system, where two cameras have uniform discretization, is studied. This study found surprisingly, that vergence movements toward an object in the 3D scene cause an increase in the depth estimation error.

In order to address this issue, we investigate two possible solutions. First we study a stereo imaging system built using cameras similar to the human eye with high resolution in the center and decreasing resolution toward the periphery. The depth estimation error in such a stereo imaging system is studied and the result is compared with the uniform resolution case. As well the optimal non-uniform resolution leading to lowest depth estimation error is studied.

Then we study a stereo imaging system built using cameras with cylindrical sensors. In this system, there is no vergence movement of the cameras, and the effect of vergence movements on depth estimation error is eliminated. The findings of this research can be used to design optimal stereo imaging systems.

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Depth Estimation Error in Stereo Imaging" submitted by Hossein Sahabi in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Co-Supervisor, Dr. Anup Basu
Department of Computing Science

Co-Supervisor, Dr. Keith Stromsmoe
Department of Electrical and Computer Engineering

Dr. Bruce Cockburn
Department of Electrical and Computer Engineering

Dr. Horacio Marquez
Department of Electrical and Computer Engineering

Dr. Mrinal Mandal
Department of Electrical and Computer Engineering

Dr. Terry Caelli
Department of Psychology

External Examiner, Dr. Minas Spetsakis.
York University, Department of Computer Science

27 July 2000

Date

Acknowledgements

I would like to express my gratitude to:

Dr. A. Basu for his encouragement, knowledge, understanding of my situation and guidance throughout this project.

Dr. Stromsmoe for his tremendous support of my thesis work. His help was extremely important in the completion of this thesis.

Dr. Reg Wilson from Carnegie Mellon University for the Tsai's calibration program.

Dr. Garry Tarr from Carleton University, Dept. of Electronics for allowing me to use their lab facilities for completing my thesis work.

Jaque Lemeiux Carleton University Dept. of Electronics System Administrator for his support for my unix account.

Mr. Stephen Gawn Senior Technical Writer at Newbridge Network Corp. for his valuable comments on the writing of this thesis.

My wife Mojgan and my son Amir for their understanding of the challenge of graduate studies.

My parents for their emotional and financial support.

To my wife Mojgan, and my son Amir.

Contents

Abstract iii

Acknowledgements iv

Dedication v

Contents vi

List of Tables x

List of Figures xii

List of Symbols xvii

Chapter 1 Introduction 1

1.1 Purpose 1

1.2 Document Overview 4

Chapter 2 Background 6

2.1 3D Structure from Stereo Imaging 6

2.1.1 Stereo Matching Algorithms 7

2.1.2 Marr-Pogio Computational Theory of Stereopsis 9

2.1.3 Grimson's Modifications of Marr-Pogio Theory 11
2.2 Error Analysis in Stereo 13
2.2.1 McVey and Lee Error Analysis 14
2.2.2 Blostein and Huang Error Analysis 15
2.2.3 Rodriguez and Aggarwal Error Analysis 16
2.2.4 Matthies and Schafer Error Analysis 17
2.2.5 Yang and Wang Error Analysis 18

Chapter 3 Vergence in Stereo Imaging **19**

3.1 Terminology and Assumptions 19
3.2 Depth of a 3D point 22
3.3 Optimal Vergence for a 3D Point 23
3.4 Optimal Vergence for an Object 30
3.5 Depth Estimation Error versus f 34
3.6 Error versus dx 35

Chapter 4 Vergence and Spatially-varying Sensing **37**

4.1 Non-uniform Discretization Model 38
4.2 Upper Bound of Depth Estimation Error 39

Chapter 5 Optimal Non-uniform Discretization **48**

5.1 Error in Estimation of X and Y Coordinates 49
5.2 Optimal Exponential Discretization 53
5.3 Optimal Linear Discretization 56
5.4 Comparison of Uniform and Non-Uniform Discretization 58

Chapter 6 Stereo Cameras with Tilt Movements **61**

6.1 Depth of A 3D Point 61

6.2 Depth Estimation Error 67

6.3 Epipolar Lines 70

Chapter 7 Spatially-Varying Stereo Cameras with Tilt Movements **74**

7.1 Upper Bound of Error in Z 74

7.2 Optimal Non-Uniform Discretization 76

Chapter 8 Stereo System with Cylindrical Sensors **79**

8.1 Cameras with Cylindrical Image Plane 80

8.2 Depth of a 3D Point 80

8.3 Depth Estimation Error Analysis 82

8.4 Comparison of Flat vs. Cylindrical Stereo 84

Chapter 9 Experimental Results **89**

9.1 Simulated Stereo Imaging System 89

9.2 Conventional Stereo Imaging System 92

9.2.1 Calibration Process 92

9.2.2 Coplanar Points 94

9.2.3 Points on a 3D Object 99

9.3 Cylindrical Stereo System 106

Chapter 10 Conclusions and Future Work **114**

References **116**

Appendix A Imaging Device **121**

Appendix B Tsai's Calibration Algorithm 124

Appendix C Stereo Images 127

List of Tables

9.1 Calibration parameters for one of the camera's positions 95

9.2 Orientation information obtained from the calibration for the three different positions 95

9.3 Depth of the selected points measured using the original images . . . 97

9.4 Depth of the selected points and the percentage error measured using the subsampled images and compared with the results obtained from the original images 98

9.5 Depth of the selected points and the percentage error measured using the non-uniform images and compared with the results obtained from the original images 98

9.6 Depth of the selected feature points using the original images 103

9.7 Depth distance between the selected feature points using the original images, and percentage error compared to the actual dimensions . . . 103

9.8 Depth of the selected feature points using the subsampled images . . 104

9.9 Depth distance between the selected feature points using the subsampled images, and percentage error compared to the actual dimensions 104

9.10 Depth of the selected feature points using the non-uniform images . . 105

9.11 Depth distance between the selected feature points using the non-uniform images, and percentage error compared to the actual dimensions 105
9.12 Stereo system parameters for depth calculation 107
9.13 Depth of the selected points measured using the stereo images 109
9.14 Depth of the selected feature points on Mr. Potato Head measured using the stereo images. 112
9.15 Depth difference of the selected feature points on Mr. Potato Head measured using the stereo images. 113

List of Figures

1.1 Depth estimation error in stereo imaging 3

2.1 McVey and Lee stereo imaging system 14

2.2 Horizontal Projection of a triangulation 15

2.3 Quantization error with normal approximation 17

3.1 Stereo imaging system 20

3.2 Stereo imaging system with vergence 21

3.3 Upper bound on the relative error in depth estimation versus the vergence angle for $P(50, Y, 300)$ 27

3.4 Variation of depth measurement uncertainty with vergence using uniform resolution images 28

3.5 Isoresolution plots for stereo cameras with uniform discretization . . . 30

3.6 Upper bound on relative error in depth estimation with non-symmetrical vergence angles for $P(25, Y, 300)$ 31

3.7 Variation of depth uncertainty for stereo cameras with uniform discretization, and non-symmetrical vergence angles 32

3.8 Isoresolution plots for stereo cameras with uniform discretization, and non-symmetrical vergence angles 32

3.9 Stereo system with an object in the scene 33
3.10 The upper bound on the average depth estimation error versus vergence angle for symmetrical movements 34
3.11 The upper bound on the average depth estimation error for $f = 50$. 35
3.12 The upper bound on average depth estimation error versus camera separation 36

4.1 Exponential pixel distribution in the image plane for $\gamma = 0.03$ 40
4.2 Upper bound of depth estimation error with non-uniform images for point $P(50, Y, 300)$ 41
4.3 Upper bound of depth estimation error with non-uniform images for point $P(25, Y, 300)$ 42
4.4 Variation of depth uncertainty with vergence using non-uniform image planes 43
4.5 Isoresolution plots for stereo cameras with non-uniform discretization 44
4.6 Average depth estimation error using non-uniform images 45
4.7 Variation of depth uncertainty for stereo cameras with non-uniform discretization and non-symmetrical vergence angles 46
4.8 Isoresolution plots for stereo cameras with uniform discretization, and non-symmetrical vergence angles 46

5.1 The error EY (left) and EZ (right) for an object using stereo cameras with non-uniform exponential discretization. 55
5.2 The error EY (left) and EZ (right) for an object using stereo cameras with non-uniform linear discretization 56
5.3 EY for using stereo cameras with uniform discretization. 59

5.4 EZ for using stereo cameras with uniform discretization for $e_x = 0.22$. 60

6.1 Stereo imaging system with tilt movements 62

6.2 Upper bound of depth estimation error of location of a 3D object in stereo imaging 71

6.3 Epipolar lines in a stereo system with vergence 73

7.1 Upper bound on depth estimation error in the location of a 3D object in stereo imaging system with non-uniform pixel distribution 77

7.2 Optimal non-uniform discretization 78

7.3 Changes in E_{min} with ν changes to keep constant overall resolution . 78

8.1 Model of the imaging system with single dimensional array 80

8.2 Stereo imaging system using cameras with cylindrical image plane . . 81

8.3 Variation of depth estimation error in stereo with vergence movements using cameras with flat image plane 85

8.4 Depth estimation error in stereo using cameras with cylindrical image plane 85

8.5 Upper bound on the depth estimation error versus the vergence angle in a stereo system with flat image plane and uniform resolution . . . 86

8.6 Upper bound of depth estimation error versus vergence angle in stereo system with flat image plane and non-uniform resolution 87

8.7 Upper bound of depth estimation error versus vergence angle in stereo system with cylindrical cameras 88

9.1 Average depth estimation error in a stereo system with uniform discretization 90

9.2 Average depth estimation error in a stereo system with non-uniform discretization. 91
9.3 Calculation of our models parameters from calibration results 94
9.4 Original stereo images in position A 96
9.5 Original stereo images in position B 96
9.6 Original stereo images in position C 96
9.7 Original stereo images position A 100
9.8 Original stereo images position B 100
9.9 Original stereo images position C 100
9.10 Original stereo images with uniform discretization, position B 101
9.11 Subsampled stereo images with uniform discretization, position B . . 101
9.12 Stereo images with non-uniform exponential discretization, position B 101
9.13 Mr. Potato Head's Dimensions (right), Mr. Potato Head's selected features (left) 102
9.14 Camera used for cylindrical stereo imaging 107
9.15 Points at which depth was estimated 108
9.16 Stereo images, right image (top) and left image (bottom), using the cameras with cylindrical image planes 110
9.17 Mr. Potato Head's Dimensions 111
9.18 Mr. Potato Head's selected features. 111

A.1 Block diagram of a CCD imaging device. 122

B.1 Tsai's camera model 124

C.1 Left original resolution image 128
C.2 Left subsampled resolution image 128

C.3 Left non-uniform resolution image 129

List of Symbols

e_x	pixel separation in horizontal direction
e_y	pixel separation in vertical direction
$P(X, Y, Z)$	3D point P with coordinates X, Y, Z
$P(\hat{X}, \hat{Y}, \hat{Z})$	Estimated values for P's coordinates
$P_R(x_r, y_r)$	Projection in the right camera
$P_R(\hat{x_r}, \hat{y_r})$	(x_r, y_r) rounded due to discretization
$P_L(x_l, y_l)$	Projection in the left camera
$P_L(\hat{x_l}, \hat{y_l})$	$(x_l, y_l$ rounded due to discretization
f	focal length
dX	Stereo cameras baseline separation
α_r	Pan vergence angle for the right camera
α_l	Pan vergence angle for the left camera
E_{min}	Min pixel separation
γ	Rate in the exponential discretization
ν	Rate in the linear discretization
α	pan vergence angle for the right camera
β	tilt angle for the right camera
γ	pan vergence angle for the left camera

θ	tilt angle for the left camera
ν	Rate of exponential discretization
R	Radius of cylindrical sensors

Chapter 1

Introduction

This chapter presents the main theme of this thesis. Section 1.1 describes the purpose of the thesis and the main idea and motivation behind this research. Section 1.2 provides an overview of the remaining chapters.

1.1 Purpose

A significant amount of research has been directed towards the development of systems that are able to perceive the three dimensional (3D) structure of objects. This 3D information is essential in many applications including geoscience, robot vision, military surveillance, autonomous navigation, and medical imaging. Stereo vision is an important method for obtaining depth information from a 3D scene. In stereo, a pair of cameras provides left and right images of a scene. The depth of each 3D point can be estimated based on the position of its projections in the two images.

Three major steps are involved in the process of deriving depth information from stereo images. First, the left and right images are pre-processed to identify well-defined features in each image. Second, correspondence is established between

features that are projections of the same physical entity in the two images. Many algorithms have been proposed to establish correspondence [1,16,17,19,21,24,25,29–31,38]. The algorithms are classified, based on the matching primitives, into area-based and feature-based techniques. Area-based techniques correspond brightness patterns in the two images [19,30]; these algorithms have well-known drawbacks [16,22]. Feature-based techniques match features such as edges [3,13,15,17,20,23,25] and linear edge segments [1,29]. In the final step of the stereopsis process, the depth of each point is obtained using triangulation.

Digital cameras have an image plane which consists of a number of discrete picture elements (pixels). In general, these pixels are uniformly arranged in a two dimensional array according to industrial standards. The projection of each 3D point in the scene is approximated to the nearest pixel. Because of this approximation the depth of each 3D point is not precise. The resulting error is known as the *discretization error.*

In stereo imaging, the discretization error generates uncertainty in estimating the depth of each 3D point. In order to model such uncertainty different approaches have been used, such as discrete tolerance limits [2,11] and multidimensional probability distribution [26]. Figure 1.1 illustrates this depth estimation error; for all the points lying in each diamond, the same depth is estimated. Therefore estimated depth of all the points lying in a diamond boundary will have an error. The upper bound of error in depth estimation is directly proportional to the size of diamonds. The diamond plots are used throughout the thesis as a tool to gain understanding of the behavior of depth estimation error upper bound. This model was first introduced by Matthies [26]. As this figure illustrates, the depth estimation error grows with distance. However, the error is not a simple function of distance — the diamonds in

the Figure 1.1 are also skewed and have different orientations.

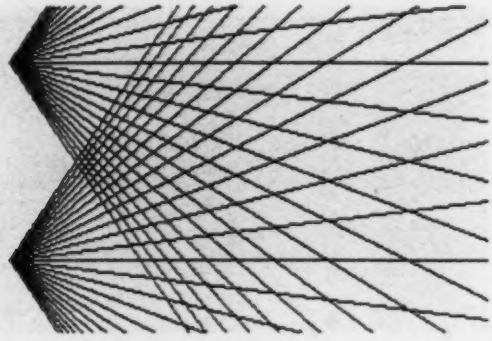


Figure 1.1: Depth estimation error in stereo imaging

In this thesis we primarily study the depth estimation uncertainty due to the discretization error. We do not consider depth estimation error due to other factors, such as matching process errors. In our study we use an active stereo imaging system. Such a system uses a dynamic pair of cameras that can be tilted (rotated about the horizontal axis) or panned (rotated about the vertical axis) independently [14, 32]. The objective of our study is to determine if the depth estimation error changes with stereo imaging system geometrical characteristics (such as vergence angle) and camera specification (such as resolution and geometrical shape of the image plane).

First we study the effect of vergence movements in a typical stereo imaging system with cameras having uniform resolution in both horizontal and vertical directions. Vergence movements of the cameras are defined as pan rotation of cameras. Cameras are typically designed with uniform resolution in both horizontal and vertical directions. Optimal vertical and horizontal resolution of stereo cameras with uniform resolution for minimizing the depth estimation error has been studied by Basu [4].

Then we present the idea of having cameras with non-uniform resolution and

examine their benefits for depth estimation error reduction. These cameras are similar to human eye, with high resolution in the center (foveal region) and decreasing resolution toward the periphery [7,35]. We also study the optimal non-uniform pixel arrangement on the camera's image plane [8].

We also explore the use of cameras with cylindrical sensors [9]. Cylindrical sensors due to the form of the image plane provide a unique feature in stereo imaging which makes vergence movements pointless. In this new approach, the negative effect of vergence movements on depth estimation error is completely eliminated.

1.2 Document Overview

Chapter 2 discusses the stereopsis process in more detail, including some well known matching algorithms. The chapter also examines the work of other researchers in the area of stereo depth estimation error.

Chapter 3 describes a typical stereo imaging system used to acquire depth information. This chapter examines the trigonometry of depth estimation and how the stereo imaging system is considered to be limited when each camera can have only horizontal (pan) vergence movements. An upper bound on depth estimation error is obtained. This upper bound is studied for variation with vergence angles of the cameras.

Chapter 4 studies a similar stereo system with cameras having non-uniform discretization in the horizontal direction and uniform resolution in vertical direction. The upper bound on the depth estimation error is studied with vergence movements of the cameras in such a stereo imaging system.

Chapter 5 describes the optimal non-uniform discretization in the horizontal

direction. Two different non-uniform discretization models (linear and exponential) are investigated.

Chapter 6 investigates a stereo imaging system that is similar to the human eye, with cameras having independent pan and tilt movements. The depth of each 3D point in the scene is formulated based on the projections of the point in the cameras as well as the system configuration. An upper bound on the depth estimation error is calculated and studied with various cameras movements.

Chapter 7 studies the stereo system described in Chapter 6 with cameras having non-uniform discretization. As well, optimal non-uniform discretization in the both horizontal and vertical directions is studied.

Chapter 8 explores a stereo imaging system that uses cameras with cylindrical sensors. The depth of each 3D point in the scene is formulated based on the coordinates of its projections in the left and right cameras. Then the upper bound on depth estimation error is calculated and studied.

Chapter 9 provides the results of experiments conducted to study the theoretical results obtained.

Chapter 10 reviews the work done and the results obtained. It discusses future work and finishes with conclusions about the project and its potential usage.

Chapter 2

Background

Stereo imaging is a widely adopted technique for obtaining 3D data. The human vision system is able to extract 3D information from two 2D images through the stereoscopic fusion of images. The process used by the human vision is not clearly understood even though much research has been devoted to developing stereo imaging systems that are able to obtain 3D information and mimic human visual system.

In this chapter, we describe some of the previous research on stereo pertaining to the thesis. Section 2.1 provides a general overview of stereo vision and how depth information is obtained from a scene using a stereopsis process. We discuss a number of different stereo matching algorithms in the literature. In section 2.2, we discuss a number of stereo error analyses which have been carried out by other researchers.

2.1 3D Structure from Stereo Imaging

In an imaging system, all the points in the viewable space are mapped into the two dimensional sensor plane of the camera. This mapping is not one-to-one; depth information can not be extracted from a single image. In order to acquire depth

information, stereo images taken from two slightly different angles are required. The depth of an object can be determined by the relative displacement between two resulting projections of the object. This displacement is referred to as the disparity between the projections.

The process of analyzing stereo images involves three main stages as follows: 1) pre-processing, 2) establishing correspondence, and 3) calculating depth.

In the preprocessing stage, the images from left and right cameras are processed independently. The objective of this step is to identify locations of some well-defined entities in two images. The entities which are identified in this step should be clearly and carefully chosen, otherwise the whole stereopsis process may lead to unsatisfactory results.

The second stage in stereopsis is establishing correspondence between images. The objective of this step is to find the entities in the two images corresponding to the same physical identity in the scene.

The third stage of stereopsis process is calculating the depth. Triangulation techniques are used to estimate depth from the disparity of matches and imaging systems geometry.

2.1.1 Stereo Matching Algorithms

There is a considerable body of computer vision literature devoted to stereo matching algorithms [16]. This section discusses only a few of the most well known techniques.

Matching algorithms are categorized based on the primitives used for matching as well as the imaging geometry. Based on the matching primitives, matching algorithms can be categorized as area-based and feature-based matching algorithms.

Area-based algorithms use brightness patterns in the vicinity of each pixel to

correlate the two images. The performance of these algorithms can be poor, because they use the intensity values at each pixel directly to correlate the two images and the intensity can vary as a result of changing viewing position and illumination.

Feature-based stereo algorithms use features in the two images, not the intensity values. They are less sensitive to changes in ambient light and perspective. Commonly used features are edge points and edge segments that can be located with subpixel precision.

Stereo matching algorithms can also be categorized based on imaging geometry. Factors that can be changed are the viewing angle of each camera and adding a third camera. In a conventional stereo system, illustrated in Figure 3.1, the optical axes of the two cameras are parallel and they are separated by a horizontal distance. We can change the imaging geometry by changing the viewing angle to have *vergence* in the two cameras. When there is vergence, greater overlap exists between the images from the two cameras.

The trinocular (three-cameras) approach to stereo matching problem has been proposed by a number of researchers, including Ohta and Kanade [31]. The third camera can be particularly useful if the focal centers of the cameras do not lie on a straight line. In this case the third camera can provide extra information to resolve ambiguities of the multiple candidate matches.

Due to the clear advantages of feature-based algorithms over area-based algorithms, most of the matching algorithms are devised as feature-based. A typical feature-based matching algorithm consists of two phases. First, a particular feature from an image is selected. Then a search is carried out in the other image to find a match. In order to be more successful in this search, various local properties of features to be matched are used. There is the possibility of having several matches

or false matches. Therefore matching will be followed by a global consistency check to eliminate multiple and/or false matches. Two constraints are used for this phase. The first is a regional disparity continuity constraint, meaning that the objects are assumed to have smooth surfaces and this smoothness leads to smooth disparities. The second constraint is figural continuity, meaning that the contours of objects in the space map to continuous curves in the images.

2.1.2 Marr-Pogio Computational Theory of Stereopsis

Marr and Pogio proposed a computational model of stereopsis based on the study of the human and other biological vision systems [25]. Marr and Pogio established two rules for matching left and right images: 1) *uniqueness*: only one disparity value is allowed to be assigned to any feature in each image. 2) *continuity*: only a small fraction of the area of an image consists of boundaries that are discontinuous in depth, and therefore the disparity in the images should vary smoothly almost everywhere.

According to this model, the human visual system uses a five-stage algorithm to solve the stereo matching problem:

1. The left and right images are spatially filtered with bar masks of four sizes and at twelve different orientations around each pixel in the images. This operation can be assumed to be approximately linear for a given intensity and contrast. These masks perform a second directional derivative after low pass filtering and smoothing and can be used to detect changes in intensity at different scales. This stage is the result of biological evidence regarding existence of independent spatial-frequency tuned channels in the human visual system.

2. For matching the left and right images, the raw data from the first stage cannot

be used because they do not directly correspond to physical features of the objects in the scene. The result of applying the filter on the image is a function with positive values in the center and negative in the skirts. Therefore zero-crossings in the filtered images resulting from the first stage are found, and these specify the position of edges in the original images.

3. For each mask size, the zero-crossings of the same sign and approximately the same orientations along corresponding horizontal lines in the two images are compared. A false match may occur in situations where zero-crossings of the same size and orientations lie very close together along horizontal lines in an image. Hence, the distance between adjacent zero-crossings of the same sign is an important parameter in this stage.

4. The human eye has some vergence movements and Marr and Pogio used biological data to make an assumption that these movements are accurately controlled by matches obtained through the various channels. To obtain fine resolution, disparity information of matches should be obtained in the smallest channels. Therefore disparity information from larger channels is used to make vergence movement in the human eye and bring the smaller channels into their disparity range.

5. When masks with certain sizes are used for the process of stereopsis, the resulting matches are stored in a temporary buffer. Two types of reasons support the existence of a memory in the human visual system. First are reasons arising from general considerations about early visual processing. Second are reasons concerning the specific problems of stereopsis. Marr and Pogio proposed the use of a dynamic memory named $2\frac{1}{2}$ *dimensional sketch*. This memory has

considerable computing power and belongs to early visual processing. The full $2\frac{1}{2}$ dimensional sketch contain approximate distances to the surfaces, as well as their orientations, contours where surface orientation changes sharply, and contours where depth is discontinuous.

2.1.3 Grimson's Modifications of Marr-Pogio Theory

The Marr-Pogio model is only a computational theory and not an explicit algorithm. Grimson designed and implemented an explicit algorithm based on the Marr-Pogio theory with particular emphasis on the matching process [17]. As well, Grimson tested the performance of the theory and an implementation on various images, and made changes to the Marr Poggio theory.

Grimson's algorithm is divided into five modules that correspond closely to the five steps in the Marr-Pogio theory. The two most important parts of Grimson's algorithm are:

- *Convolution:* The left and right images must be filtered to transform them into a form that the matcher may operate. Marr and Pogio proposed convolving images at twelve different orientations with special bar masks for which the cross section was the difference of two gaussians. Grimson used the Marr-Hildreth approach, and found that the intensity changes that occur at all orientations may be detected by using a single non-oriented Laplacian operator. Therefore the left and right images are convolved with a series of two-dimensional operators in the form of a Laplacian of a Gaussian or by an approximation to this operator.

- *Matching:* The matching process proposed by Grimson proceeds in a coarse to

fine manner. The algorithm starts with images filtered by the largest filters, and by reducing the resolution of points greatly reduces the search space and makes matching easier. The matches obtained from coarser details are used to restrict the matching in finer details. For each size of filter, matching consists of six steps:

1. Fix the eye position and get the right and left images.
2. Locate a zero-crossing in one image.
3. Partition the region around the same location in the other image into three pools or subregions consisting of two convergent and divergent pools and a smaller pool between them. These pools are the areas to be searched for matching zero-crossings.
4. Assign a match to the selected zero-crossing that come from convolutions with the same size filter, and of the same sign and orientation. If exactly one match with the above conditions is found in a pool, the location of that match is transformed to the matcher. If two matches are found within the same pool, the matcher is informed and no match will be assigned for the point. If just one match in only one pool is located, that match is accepted and the disparity is recorded in the buffer.
5. If multiple candidates for matching were obtained in step 4, resolve the ambiguity in matching. This is carried out by scanning the neighborhood around the point which is considered and studying the sign of disparity (sign of the disparity is the sign of pool that the match comes from, such as convergent, divergent, and zero) of ambiguous matches within that neighborhood. Then the right match is selected as the one with the same

disparity sign.

6. Store the disparity value in a buffer.

Grimson made modifications to the Marr-Pogio theory [18]. Two of these modifications are:

- *Continuity:* Grimson's implementation uses a figural continuity in disparity to check the regions to validate the matches. Because surfaces are smooth, there should be a continuity in disparity; however, Grimson observed that most matching errors occur along discontinuities in depth, for instance occluding objects on the scene. He proposed to impose continuity in disparity along the contours in the image, rather than disparity continuity over an area of the image. This is the figural continuity constraint proposed by Mayhew and Frisby [27].

- *Vertical Disparity:* In the Marr-Pogio theory, searches for matches are carried out along the same horizontal line. However, local distortions due to perspective effects of noise and discretization error cause difficulty in finer resolutions. Therefore Grimson proposed searching for matching points in two images, not only along the same horizontal line, but also along a number of lines above and below the horizontal line.

2.2 Error Analysis in Stereo

The depth calculation stage in stereopsis relies on the location of matches found in the matching stage. Discretization error as a result of discrete pixels in the camera's image plane is the major source of error in stereo imaging. In this section, we discuss the work of other researchers' in modeling this error.

2.2.1 McVey and Lee Error Analysis

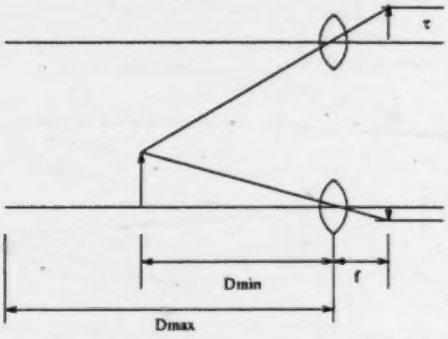


Figure 2.1: McVey and Lee stereo imaging system

McVey and Lee [28] have performed a worst case error analysis on the image plane resolution required to achieve depth estimation of a given accuracy. Figure 2.1 illustrates the stereo imaging system that they used for their analysis. McVey and Lee simplified their analysis by assuming that the upper image can be approximated by the shifted lower image and the shift τ can be obtained by appropriate signal processing techniques. According to their analysis, the depth measurement error can be obtained as follows:

$$e = \frac{\left|\tilde{D} - D\right|}{D} = \frac{100 l_p}{fL} D \tag{2.1}$$

where l_p is the length of each pixel, f is the focal length, L is the separation of the two cameras, D is the depth, and 100 is for the percentage error. McVey and Lee then used their analysis of the resulting error to provide a synthesis approach to determine stereo imaging system parameters from a number of specifications, including maximum absolute distance measurement error, and distance extremes D_{min} and D_{max}.

2.2.2 Blostein and Huang Error Analysis

A number of other researchers considered another alternative to the worst case error analysis, which would be to determine the probability that a certain position estimate is within a specified position tolerance given the specifications of the stereo imaging system. Blostein and Huang [10] performed an error analysis in a stereo imaging system with parallel cameras as illustrated in Figure 2.2. In this figure $P_1P_2P_3P_4$ defines the region of uncertainty.

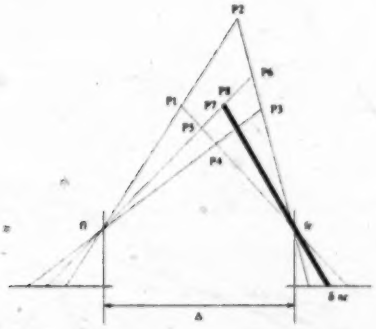


Figure 2.2: Horizontal Projection of a triangulation

Let us assume 3D point S has projections (I_l, J_l) and (I_r, J_r) in the left and right cameras, respectively. For a parallel stereo imaging system $I_l = I_r$. The focal length of the cameras is denoted by f, the baseline of two cameras is denoted by Δ, and the pixel dimension is represented by dv. The following equations provide an approximation to the length of the segment $\overline{P_5P_6}$

$$\|P_5 - P_6\| \approx \frac{\Delta}{(J_l - J_r + n_l)(J_l - J_r + n_l - 1)}[\frac{f^2}{dv^2} + J_l^2]^{1/2} \tag{2.2}$$

Similarly the following equation provides the length of the segment $\overline{P_7P_8}$.

$$\|P_7 - P_8\| \approx \frac{\Delta dn_r}{(J_l - J_r + n_l - n_r)^2}[\frac{f^2}{dv^2} + J_l^2]^{1/2} \tag{2.3}$$

In these equations n_l and n_r are real numbers between 0 and 1 representing the quantization uncertainty.

Assuming a point is uniformly distributed in the region of uncertainty, the joint density of the distribution on the two image planes is obtained from the following equation:

$$f_{N_l,N_r}(n_l, n_r) = \frac{1}{-\ln(|1-(J_r-J_l)^{-2}|(J_l-J_r+n_l-n_r)^2} \tag{2.4}$$

where N_l and N_r are the random variables representing probablistic interpretation of n_l and n_r. From this distribution function, the probabilities of error in three dimensions X, Y and Z are less than a certain limit τ_z can be obtained. For the depth (Z component) we have:

$$P(\epsilon_z < \tau_z) = \left\{ \begin{array}{ll} 1-(1-\tau_z D)^2 & \tau_z < \frac{1}{D} \\ 1 & \tau_z >= \frac{1}{D} \end{array} \right\} \tag{2.5}$$

where D represents the horizontal disparity between left and right image.

2.2.3 Rodriguez and Aggarwal Error Analysis

Rodriguez and Aggarwal [34] carried out a stochastic analysis of the depth estimation error. They calculated the expected value of the absolute depth estimation error ΔZ as follows:

$$E[|\Delta Z|] \approx \frac{\delta}{9bf}[z_{min}^2 + z_{min}z_{max} + z_{max}^2] \tag{2.6}$$

When measuring error, the relative depth estimation error is more meaningful than the absolute error. Rodriguez and Aggarwal defined the relative depth estimation error ϵ as follows:

$$\epsilon = \frac{|\Delta Z|}{z_{max} - z_{min}} \tag{2.7}$$

where z_{max} and z_{min} define the range of the Z coordinate. They believe that ϵ describes the depth resolution better than the percentage error $|\Delta Z|/Z$ does. The expected value of ϵ is given by:

$$E[\epsilon] = \frac{E(|\Delta Z|)}{z_{max} - z_{min}} \tag{2.8}$$

2.2.4 Matthies and Schafer Error Analysis

Matthies and Schafer [26] used a 3D gaussian distribution to model the quantization error in stereo imaging. They assumed that the error in the measured image coordinates is normally distributed (*i.e.* Gaussian). Then they defined a 3D gaussian distribution for 3D coordinates. For the 3D coordinates, the true distribution will not be Gaussian because of the non-linear nature of the triangulation process. Figure 2.3 illustrates the normal approximation of the error. In the figure, the ellipse represents the contour of the error model and the diamond represents the actual quantization error. For nearby points, the contour will be close to spherical and therefore the approximation is adequate; for points further away, the more eccentric they become, the approximation becomes less accurate.

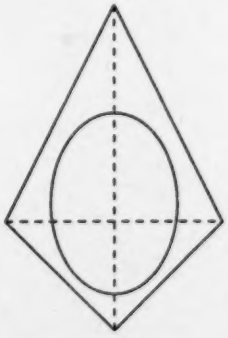


Figure 2.3: Quantization error with normal approximation

2.2.5 Yang and Wang Error Analysis

Yang and Wang [39] have studied sources of errors in 3D shape reconstruction from structured lighting. In the structured lighting technique, a spatially modulated pattern (typically a grid pattern) is projected onto the scene to encode the image object for analysis. Yang and Wang identified three type of errors: system modeling error, image processing error, and experimental error. System modeling error is the error introduced in the calculations due to cameras not behaving according to an ideal parallel projection model where projection rays make an angle with respect to the optical access. Image processing errors which are categorized according to quantization error and error due to mislocating features in the image plane. Finally, experimental errors are errors due to calibration and operation of the imaging equipments. Yang and Wang studied the error in inferring 3D surface orientation and analyzed the principle surface curvature. Upper bounds on the error were established in each case.

Chapter 3

Vergence in Stereo Imaging

In this chapter we study the effects of vergence in stereo imaging on depth measurement error. Section 3.1 introduces the terminology and assumptions that we have adopted in our analysis. These assumptions, as well as terms, are generally valid through the whole thesis unless stated otherwise. Section 3.2 presents the depth of a 3D point as a function of the coordinates of its projections in the left and right cameras and the vergence angle. In section 3.3 we study the optimal vergence angle in the stereo system leading to minimal depth estimation error, if there is any. Section 3.4 discusses how the depth estimation error behaves with changes in the focal length of the cameras, as well as their baseline separation. Section 3.5 presents the depth estimation error behavior with changes in focal length. Section 3.6 discusses the depth estimation error behavior with changes in dX.

3.1 Terminology and Assumptions

In this section, we define a number of terms that are used for our analysis. Figure 3.1 illustrates the configuration of a simple stereo imaging system, where two cameras

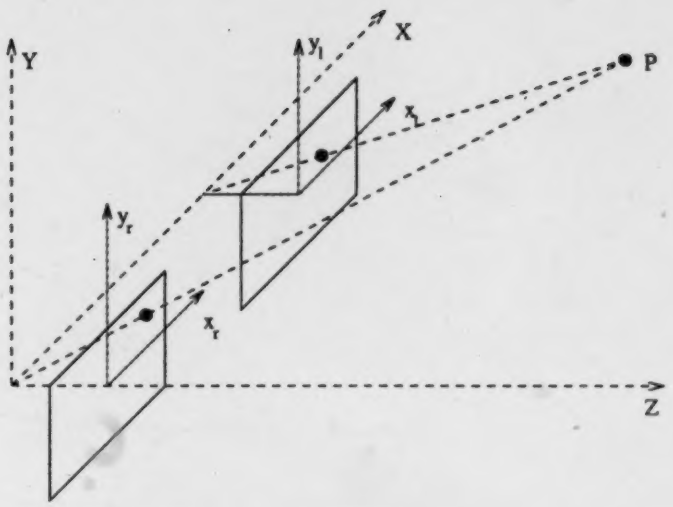


Figure 3.1: Stereo imaging system

have parallel image planes (no vergence) and are merely separated in the X direction. The focal point of each camera is in fact in front of its image plane; we use the symmetry of the real focal point with respect to the image plane to simplify our analysis. Assuming that the two cameras have the same focal length, the focal length of each camera is denoted by f. The separation distance between the cameras is denoted by dX. It is assumed that the origin of the 3D world coordinates is at the focal point of the right camera. Each camera has a uniform pixel arrangement in both vertical and horizontal directions. The distance between two adjacent pixels along the x-direction is denoted by e_x, and similarly the distance between two adjacent pixels along the y-direction is denoted by e_y. The projection of the 3D point $P(X, Y, Z)$ in the right and left cameras is represented by (x_r, y_r) and (x_l, y_l), respectively. Because of the discrete placement of pixels, these two projections are approximated by $(\hat{x}_r, \hat{y}_r)$ and $(\hat{x}_l, \hat{y}_l)$. The discretization error in turn leads to an estimate $(\hat{X}, \hat{Y}, \hat{Z})$ of the coordinates of point P.

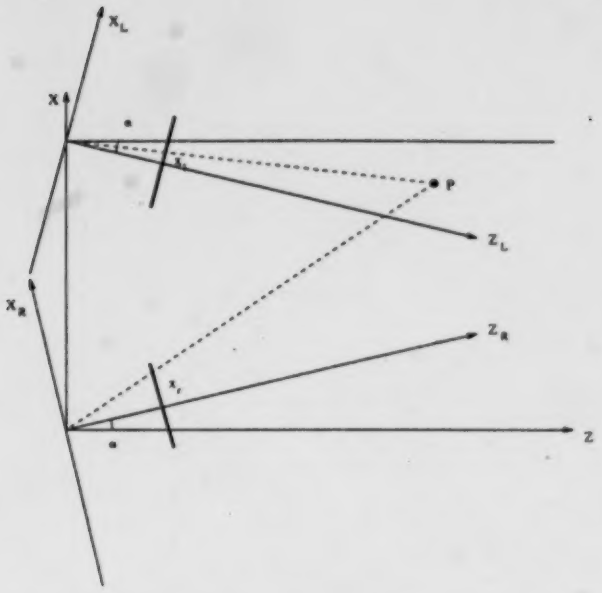


Figure 3.2: Stereo imaging system with vergence

The major assumption throughout this thesis is the only source of error is image plane quantization. A pinhole camera model is used and we ignore the lens distortion and the other forms of lens imperfections. In the pinhole camera model, lens is considered to be a point through which all incoming rays of light passes. In reality a lens with a finite aperture size is used, and the finite aperture size allows only one range to be focused on. However it is a valid assumption that under well lighted conditions and with proper geometry setup the aperture size is small enough such that the effect of blurring is small and the pinhole model holds across the scene. In addition to that in our analysis throughout this thesis we assume matching between points in the left and right cameras has already been performed and correct matches have been identified without any error.

3.2 Depth of a 3D point

Figure 3.2 illustrates a two-dimensional view of a stereo system along the Y axis toward the origin. The left and right cameras have vergence angles α_L and α_R, respectively. In this chapter we ignore the noise and the effect of Y coordinate in the depth estimation error. Theorem 1 provides a formula for the calculation of depth (Z coordinate) of the point P in Figure 3.2.

Theorem 1 *The depth of point P is given by the equation below. For definition of parameters refer to Section 3.1 and List of Symbols.*

$$\begin{aligned} Z = dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)/ \\ ((f\cos\alpha_L + x_l\sin\alpha_L)(f\sin\alpha_R + x_r\cos\alpha_R) \\ +(f\sin\alpha_L - x_l\cos\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)) \end{aligned} \tag{3.1}$$

Proof: In order to prove this theorem, a new coordinate system for each camera is defined, as illustrated in Figure 3.2. The origin of the new coordinate system is located in the focal point of the camera and its Z axis is perpendicular to the image plane. Let (X_R, Y_R, Z_R) and (X_L, Y_L, Z_L) be the coordinates of P in the new coordinate systems, which correspond to the right and left cameras, respectively. The following relations exist between the projection of P and its coordinates in each of these systems:

$$x_r = \frac{fX_R}{Z_R} \qquad x_l = \frac{fX_L}{Z_L} \tag{3.2}$$

X_R, Z_R, X_L and Z_L can be converted to X and Z (the main 3D world coordinate system) using the following equations:

$$Z_R = Z\cos\alpha_R + X\sin\alpha_R \qquad X_R = X\cos\alpha_R - Z\sin\alpha_R \tag{3.3}$$

$$Z_L = Z\cos\alpha_L - (X - dX)\sin\alpha_L \qquad X_L = (X - dX)\cos\alpha_L + Z\sin\alpha_L \tag{3.4}$$

Note that α_L is the rotation of the main 3D coordinate system in the clockwise direction, and α_R is the rotation of main 3D coordinate system in the counterclockwise direction. Using these relationships, Equation 3.2 is rewritten as:

$$x_r = \frac{f(X\cos\alpha_R - Z\sin\alpha_R)}{(Z\cos\alpha_R + X\sin\alpha_R)} \qquad x_l = \frac{f((X - dX)\cos\alpha_L + Z\sin\alpha_L))}{(Z\cos\alpha_L - (X - dX)\sin\alpha_L)} \tag{3.5}$$

By eliminating X from these equations, equation 3.1 is obtained.

It should be noted that this theorem does not consider the size of the image plane of the cameras. In reality the size of the image plane is limited and, for certain vergence angles, point P may not lie in the field of view of the cameras.

3.3 Optimal Vergence for a 3D Point

In order to study the effect of vergence on the depth estimation error for a single 3D point, we would like to obtain an upper bound on the maximum error in depth estimation. Theorem 2 provides such an upper bound.

Theorem 2 *An upper bound on the relative error in depth of point $P(X, Y, Z)$ is given by the equation below. For definition of parameters refer to Section 3.1 and*

List of Symbols.

$$E_Z = \left| \frac{\hat{Z} - Z}{Z} \right| \leq \frac{e_x \sin\alpha_L}{2f} \frac{Z\cos\alpha_L - (X - dX)\sin\alpha_L}{Z} + \frac{e_x \sin\alpha_R}{2f} \frac{Z\cos\alpha_R + X\sin\alpha_R}{Z} + \frac{e_x \sin\alpha_L}{2fdX} \frac{X(Z\cos\alpha_L - (X - dX)\sin\alpha_L)}{Z} + \frac{e_x \cos\alpha_R}{2fdX}(Z\cos\alpha_R + X\sin\alpha_R) + \frac{e_x \cos\alpha_L}{2fdX}(Z\cos\alpha_L - (X - dX)\sin\alpha_L) - \frac{e_x \sin\alpha_R}{2fdX} \frac{(X - dX)(Z\cos\alpha_R + X\sin\alpha_R)}{Z} \tag{3.6}$$

Proof: The projection of point P in each camera has at most $e_x/2$ discretization error. We have:

$$\hat{x}_r = x_r \pm e_x/2 \qquad \hat{x}_l = x_l \pm e_x/2 \tag{3.7}$$

Using Theorem 1, the depth of point P is estimated as:

$$\hat{Z} = dX(f\cos\alpha_L + \hat{x}_l \sin\alpha_L)(f\cos\alpha_R - \hat{x}_r \sin\alpha_R)/ ((f\cos\alpha_L + \hat{x}_l \sin\alpha_L)(f\sin\alpha_R + \hat{x}_r \cos\alpha_R) + (f\sin\alpha_L - \hat{x}_l \cos\alpha_L)(f\cos\alpha_R - \hat{x}_r \sin\alpha_R)) \tag{3.8}$$

By substituting for $\hat{x}_r$ and $\hat{x}_l$ in the above equation, we have:

$$\begin{aligned} \hat{Z} &= dX(f\cos\alpha_L + x_l \sin\alpha_L \pm \frac{e_x \sin\alpha_L}{2})(f\cos\alpha_R - x_r \sin\alpha_R \pm \frac{e_x \sin\alpha_R}{2})/ \\ &[(f\cos\alpha_L + x_l \sin\alpha_L \pm \frac{e_x \sin\alpha_L}{2})(f\sin\alpha_R + x_r \cos\alpha_R \pm \frac{e_x \cos\alpha_R}{2}) + \\ &(f\sin\alpha_L - x_l \cos\alpha_L \mp \frac{e_x \cos\alpha_L}{2})(f\cos\alpha_R - x_r \sin\alpha_R \mp \frac{e_x \sin\alpha_R}{2})] \\ &= Num(Z)(1 \pm \frac{e_x \sin\alpha_L}{2(f\cos\alpha_L + x_l \sin\alpha_L)})(1 \pm \frac{e_x \sin\alpha_R}{2(f\cos\alpha_R - x_r \sin\alpha_R)})/ \\ &\qquad (Den(Z) \pm G(Z)) \end{aligned} \tag{3.9}$$

where $Num(Z)$ and $Den(Z)$ represent the numerator and denominator of the right hand side of equation 3.1, and $G(Z)$ is:

$$G(Z) = \frac{e_x \sin\alpha_R}{2}(f\sin\alpha_R + x_r\cos\alpha_R) + \frac{e_x\cos\alpha_L}{2}(f\cos\alpha_L + x_l\sin\alpha_L) + \frac{e_x\cos\alpha_R}{2}(f\cos\alpha_R - x_r\sin\alpha_R) + \frac{e_x\sin\alpha_L}{2}(f\sin\alpha_L - x_l\cos\alpha_L) \quad (3.10)$$

By replacing $Z/Num(Z)$ for $Den(Z)$, Equation 3.9 is simplified to:

$$\hat{Z} = Z(1 \pm \frac{e_x\sin\alpha_L}{2(f\cos\alpha_L + x_l\sin\alpha_L)})(1 \pm \frac{e_x\sin\alpha_R}{2(f\cos\alpha_R - x_r\sin\alpha_R)})/ (1 \pm Z\frac{G(Z)}{dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)}) \quad (3.11)$$

For many practical situations, such as a room environment where Z is not very large, $Z\frac{G(Z)}{dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)}$ is small and the above equation can be rewritten into the following form:

$$\hat{Z} \cong Z(1 \pm \frac{e_x\sin\alpha_L}{2(f\cos\alpha + x_l\sin\alpha_L)})(1 \pm \frac{e_x\sin\alpha_R}{2(f\cos\alpha_R - x_r\sin\alpha_R)}) (1 \mp Z\frac{G(Z)}{dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)}) \quad (3.12)$$

By expanding the right hand side of the above equation, some of the terms that are the result of multiplying two or three fractional terms in the parentheses are negligible. Thus:

$$\hat{Z} \cong Z(1 \pm \frac{e_x\sin\alpha_L}{2(f\cos\alpha_L + x_l\sin\alpha_L)} \pm \frac{e_x\sin\alpha_R}{2(f\cos\alpha_R - x_r\sin\alpha_R)} \mp Z\frac{G(Z)}{dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)}) \quad (3.13)$$

By substituting $G(Z)$ from Equation 3.10 the worst case relative error of Z is given by:

$$|\frac{\hat{Z}-Z}{Z}| \leq \{\frac{e_x \sin\alpha_L}{2(f\cos\alpha_L + x_l \sin\alpha_L)} + \frac{e_x \sin\alpha_R}{2(f\cos\alpha_R - x_r \sin\alpha_R)} + \tag{3.14}$$
$$\frac{e_x \sin\alpha_R Z(f\sin\alpha_R + x_r\cos\alpha_R)}{2dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)} + \frac{e_x\cos\alpha_R Z}{2dX(f\cos\alpha_R - x_r\sin\alpha_R)} +$$
$$\frac{e_x\cos\alpha_L Z}{2dX(f\cos\alpha_L + x_l\sin\alpha_L)} + \frac{e_x\sin\alpha_L Z(f\sin\alpha_L - x_l\cos\alpha_L)}{2dX(f\cos\alpha_L + x_l\sin\alpha_L)(f\cos\alpha_R - x_r\sin\alpha_R)}\}$$

Using Equation 3.5, the above equation is rewritten as:

$$E_Z = |\frac{\hat{Z}-Z}{Z}| \leq \frac{e_x\sin\alpha_L}{2f}\frac{Z\cos\alpha_L - (X-dX)\sin\alpha_L}{Z} +$$
$$\frac{e_x\sin\alpha_R}{2f}\frac{Z\cos\alpha_R + X\sin\alpha_R}{Z} +$$
$$\frac{e_x\sin\alpha_L}{2fdX}\frac{X(Z\cos\alpha_L - (X-dX)\sin\alpha_L)}{Z} +$$
$$\frac{e_x\cos\alpha_R}{2fdX}(Z\cos\alpha_R + X\sin\alpha_R) +$$
$$\frac{e_x\cos\alpha_L}{2fdX}(Z\cos\alpha_L - (X-dX)\sin\alpha_L) -$$
$$\frac{e_x\sin\alpha_R}{2fdX}\frac{(X-dX)(Z\cos\alpha_R + X\sin\alpha_R)}{Z} \tag{3.15}$$

According to this theorem, the upper bound on relative error is directly proportional to e_x, as was expected. In other words, the higher the resolution of the cameras, the lower the error. Furthermore, the upper bound of error is inversely proportional to the focal length. Therefore, if the cameras zoom into a point, the error is likely to be reduced.

Now let us consider the behavior of the error upper bound with respect to changes in the vergence angle. For this purpose we use a numeric example. Suppose f is 10 mm, the size of image plane is 40 mm × 40 mm, dX is 100 mm, and e_x is 0.05 mm. Figure 3.3 illustrates the changes in the upper bound of relative depth

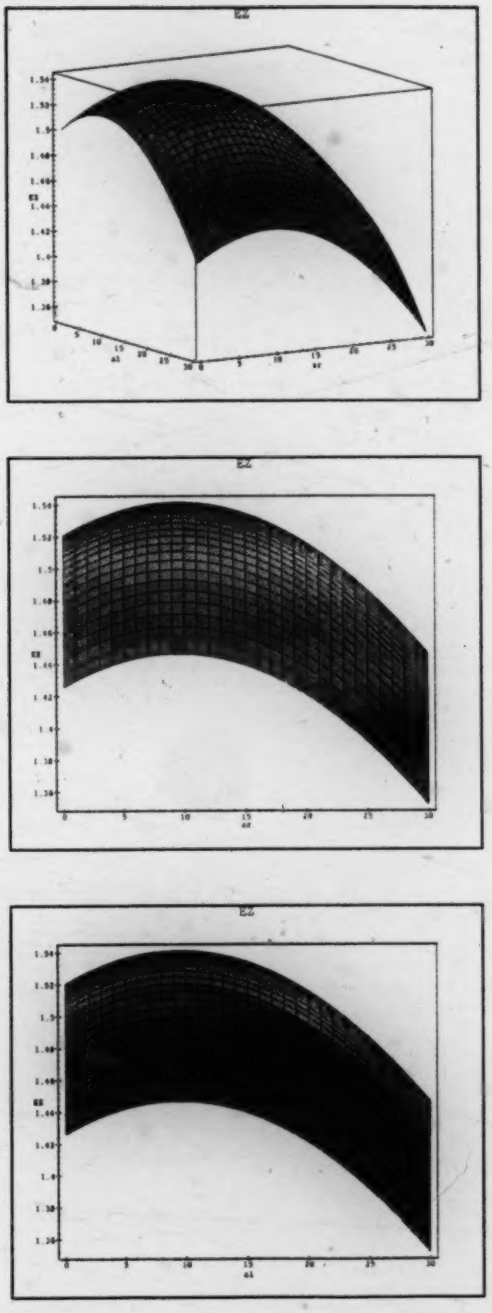


Figure 3.3: Upper bound on the relative error in depth estimation versus the vergence angle for $P(50, Y, 300)$

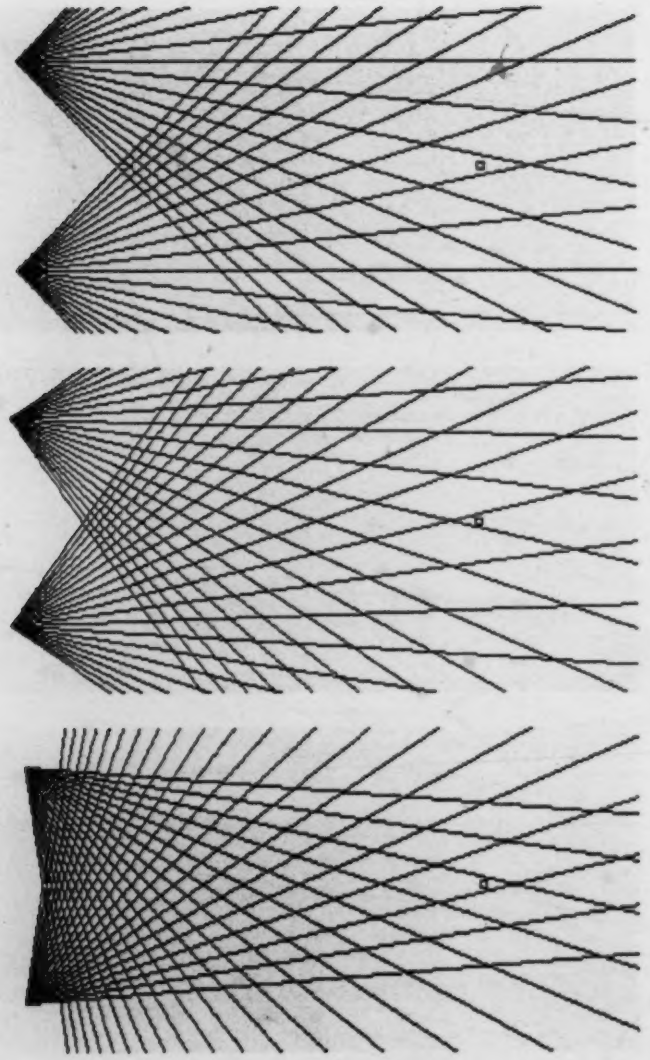


Figure 3.4: Variation of depth measurement uncertainty with vergence using uniform resolution images

estimation error for a point P located at $X = 50$ mm and $Z = 300$ mm , with changes in α_L and α_R. Point P in this example is situated symmetrically with respect to the two cameras. The error upper bound increases with vergence and it is at its maximum when both α_L and α_R are roughly 10 degrees. The results illustrated in Figure 3.3 are remarkable due to the fact that, unlike common perception, vergence likely increases the depth estimation error.

Figure 3.4 illustrates the variation of depth uncertainty of a point P for three different vergence angles graphically; the size of the diamond that covers the point P changes according to the plot of Figure 3.3. Figure 3.4 demonstrates that the maximum depth estimation error occurs when the projection of the point P is at the center of the image plane of both cameras.

Lemma 1 *Suppose point P is located symmetrically with respect to two cameras. The vergence angle corresponding to the maximum error for this point is obtained from the equation below. For definition of parameters refer to Section 3.1 and List of Symbols.*

$$\alpha_{max} = \arctan(\frac{dX/2}{Z}) \tag{3.16}$$

Using this lemma with the system in the above example, we have:

$$\begin{aligned} \alpha_{max} &= \arctan(\frac{100/2}{300}) \\ &= 9.46 \end{aligned} \tag{3.17}$$

which is consistent with the graph in Figure 3.3.

Figure 3.5 gives isoresolution plots for our stereo imaging system. Each curve represents the points in the scene that have the same upper bound value for error in depth estimation. For the zero vergence angle, the isoresolution plots for the points

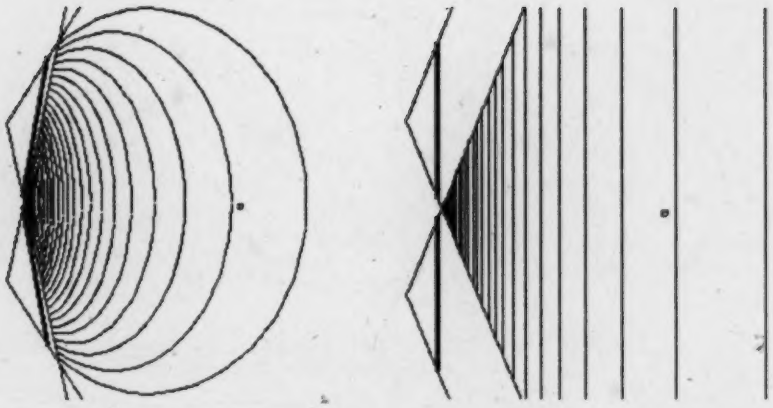


Figure 3.5: Isoresolution plots for stereo cameras with uniform discretization

whose depths are not very large compared to dx are almost straight lines. When the vergence angle is increased, the plots become curved. The resolution of the cameras directly affects the distance between the curves in the isoresolution plots. The higher the resolution of the cameras, the denser the isoresolution plots. The isoresolution curve's density around the fixed point P in the scene changes depending on the vergence angle - the density of the curves around P increases when the cameras turn toward point P. Figure 3.6 illustrates the variation of depth estimation error when point P is not located symmetrically with respect to two cameras. The upper bound of error in depth estimation is maximum when projection of point P in both cameras is in the center of the image plane. The result is also graphically confirmed in Figure 3.7 (diamond plots) and in Figure 3.8 (isoresolution plots).

3.4 Optimal Vergence for an Object

The upper bound on the relative error in depth estimation that we obtained in the previous section may have little practical value. Due to the randomness of the position

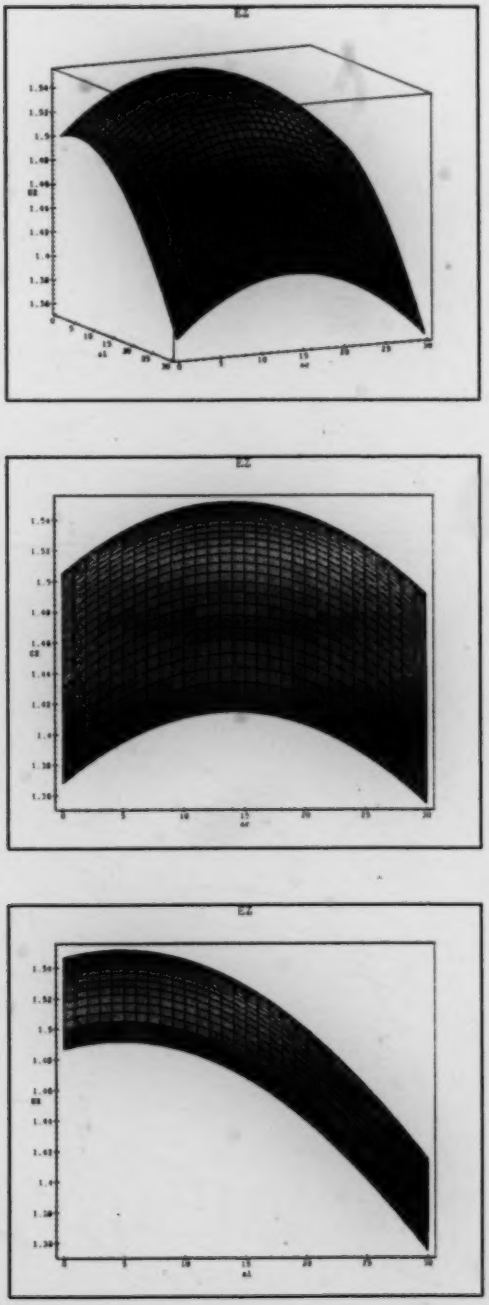


Figure 3.6: Upper bound on relative error in depth estimation with non-symmetrical vergence angles for $P(25, Y, 300)$

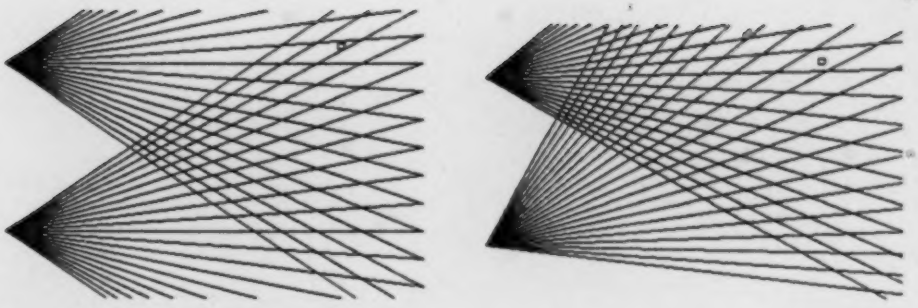


Figure 3.7: Variation of depth uncertainty for stereo cameras with uniform discretization, and non-symmetrical vergence angles

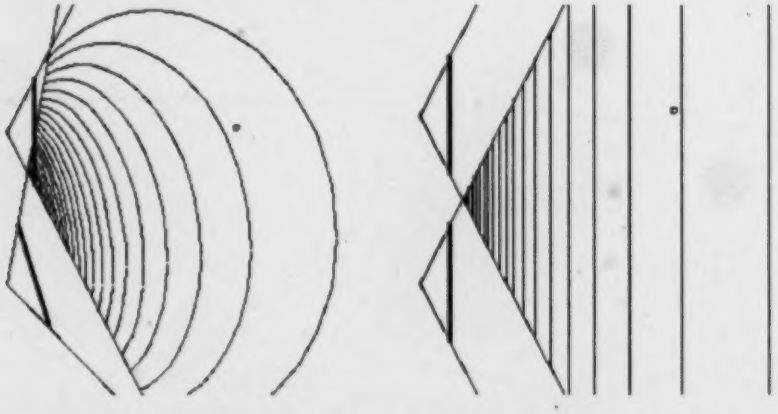


Figure 3.8: Isoresolution plots for stereo cameras with uniform discretization, and non-symmetrical vergence angles

of an arbitrary 3D point in the scene, the actual error is not necessarily equal to the upper bound. In this section we use our upper bound for a single point to obtain upper bound on the average error in depth estimation for all the points lying in a region in the scene. This is especially important when an object of interest in the scene with finite dimensions is considered.

Figure 3.9 illustrates a region in the scene representing the bounding box of an arbitrary object in the scene defined by the following constraints:

$$Zmin < Z < Zmax \qquad Xmin < X < Xmax$$

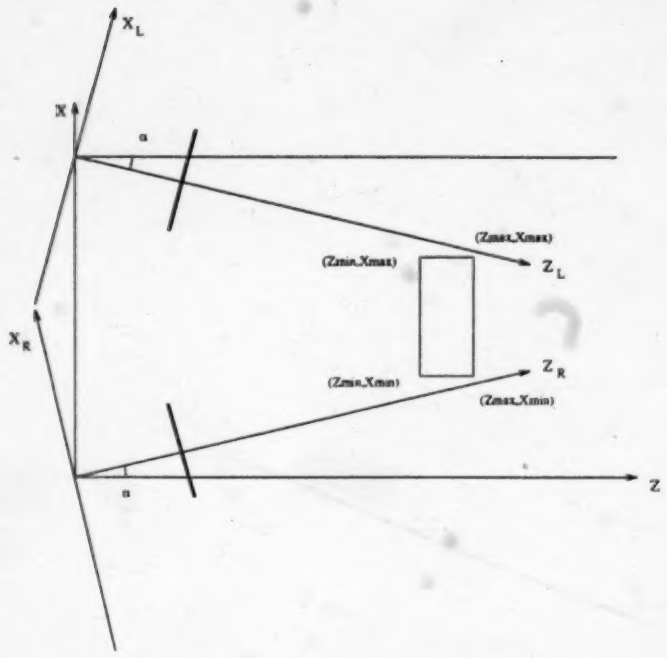


Figure 3.9: Stereo system with an object in the scene

The average depth estimation error for all of the points lying in this area is calculated as follows:

$$\overline{E_Z} = \frac{\int_{Xmin}^{Xmax} \int_{Zmin}^{Zmax} E_Z dX dZ}{\int_{Xmin}^{Xmax} \int_{Zmin}^{Zmax} dX dZ} \tag{3.18}$$

Due to the complexity of the upper bound of the E_Z equation, we use numerical methods for obtaining $\overline{E_Z}$.

Suppose an object of interest (in a system with $f = 10$ mm, 40 mm $\times$ 40 mm imaging area, $e_x = 0.05$ mm and $dX = 100$ mm) is located in the region of $280 < Z < 320$ and $35 < X < 65$. For simplicity we assume that α_L and α_R are equal. Figure 3.10 shows the average error for this object as a function of vergence angle. The upper bound is at a maximum when the center of the object region is projected onto the centers of the cameras. For small objects the upper bound on the

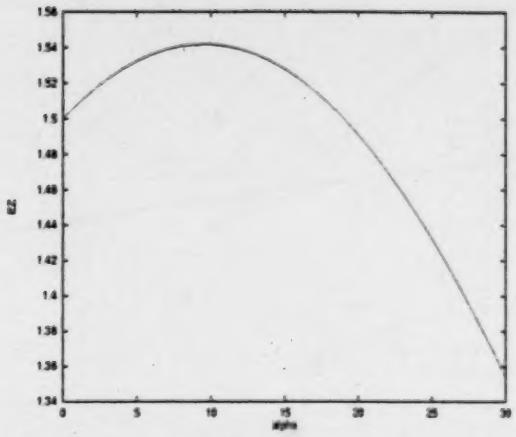


Figure 3.10: The upper bound on the average depth estimation error versus vergence angle for symmetrical movements

average error is close to the upper bound on the error for the center of the object.

3.5 Depth Estimation Error versus f

The results from the previous sections indicate that the average depth estimation error is inversely proportional to the focal length f. Therefore we attempt to reduce the depth estimation error by altering both the vergence and focal lengths. Figure 3.11 plots the upper bound on the average depth estimation error of the previous section object when both cameras' focal length is increased to 50 mm; the upper bound is sharply decreased compared to f=10 mm. By increasing the focal length the projection of the object covers a greater area of the image plane. The minimum upper bound on depth estimation error is attained when the projection of the object covers the whole image plane of the cameras. Such a situation for an object placed symmetrically with respect to two cameras occurs for unique values of focal length and vergence angle.

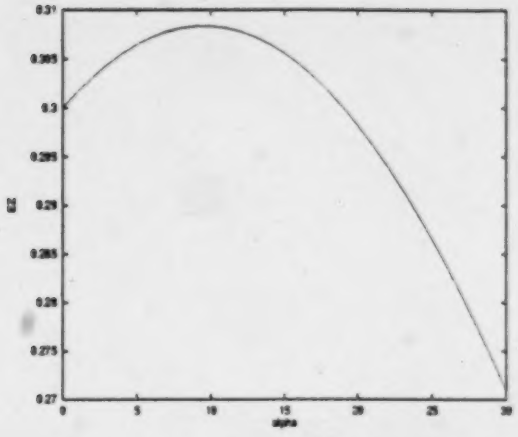


Figure 3.11: The upper bound on the average depth estimation error for $f = 50$

3.6 Error versus dx

The upper bound on depth estimation error is also dependent on the stereo cameras' separation dx. Figure 3.12 illustrates the behavior of the upper bound with changes in dx. In order to draw this plot for each value of dx, both cameras' vergence angles were changed to keep both cameras panned toward the object at all times; for each new configuration based on the value of dX the peak upper bound is always obtained. As Figure 3.12 illustrates, the upper bound decreases with wider separations between two cameras. Note that $dX = 0$ illustrated in the figure is obviously meaningless.

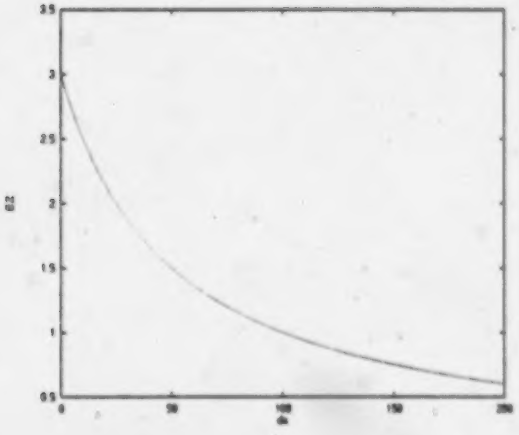


Figure 3.12: The upper bound on average depth estimation error versus camera separation

Chapter 4

Vergence and Spatially-varying Sensing

In Chapter 3 we analyzed the effect of vergence on depth estimation error in a stereo imaging system with cameras having uniform discretization. With uniform discretization, the depth estimation error can be reduced but the minimum upper bound only occurs when the object of interest is in the edge of the image. In other words, there is no optimal vergence angle which leads to minimum error while maintaining a good coverage of the background scene.

In order to solve this problem, we investigate the use of cameras with non-uniform discretization. In fact, we investigate a stereo imaging system similar to the human visual system. In the human eye, there is a high resolution foveal region in the center and the resolution decreases towards the periphery.

Spatially-varying sensing is an important technique for reducing the processing required on images. For instance in the human visual system due to high resolution, a large amount of processing is required for the foveal region. If the human visual system

resolution was uniform, the brain processing power must have been considerably higher to cover all the visual field.

Several approaches for modeling variable resolution images have been proposed [5, 12, 36]. In this chapter we present two different models of non-uniform discretization: linear and exponential. It is assumed that our cameras have a uniform pixel arrangement in the vertical (Y) direction and non-uniform linear or exponential pixel arrangement in the horizontal (X) direction. Section 4.1 describes the non-uniform discretization models. Section 4.2 explores the upper bound of depth estimation error for a point and a 3D object.

4.1 Non-uniform Discretization Model

We consider two models of non-uniform discretization: exponential and linear. We only study the exponential model in depth; the linear model is skipped because it is very similar.

1. *Exponential Discretization:* The non-uniform exponential pixel arrangement is essentially based on the fish-eye transform [6]. The fish-eye transform, which is based on the characteristics of fish-eye lenses, describes a variable resolution mapping of a uniform resolution image to an image with high resolution in the center and non-linearly decreasing resolution towards the periphery. Based on this transform, any point with coordinates $(\hat{x}, \hat{y})$ in the variable resolution plane

is mapped to the point (x, y) in the uniform resolution plane as follows:

$$\rho = \sqrt{\hat{x}^2 + \hat{y}^2} \qquad \hat{\theta} = \arctan(\frac{\hat{y}}{\hat{x}})$$

$$r = \frac{e^{\frac{\rho}{s}} - 1}{\lambda} \qquad \hat{\theta} = \theta$$

$$x = r\cos\theta \qquad y = r\sin\theta \tag{4.1}$$

In the above equations, s is a simple scaling factor and λ controls the amount of distortion over the entire range.

Based on the idea of the fish-eye transform, we define the pixel separation as:

$$e_x(i) = E_{min}e^{\gamma i} \tag{4.2}$$

where E_{min} is the smallest pixel separation in the center of the camera, γ is the factor that determines the rate of increase in pixel separation, and i is the pixel count in the X direction from the center of image plane. Figure 4.1 illustrates the pixel distribution for $\gamma = 0.03$. Note that the pixel separation is merely non-uniform in the X direction.

2. *Linear Discretization:* The non-uniform linear discretization model is based on the following equation:

$$e_x(i) = E_{min} + \nu i \tag{4.3}$$

where E_{min} is the pixel separation in the center of the image plane and ν is a constant coefficient which controls the rate of increasing pixel separation.

4.2 Upper Bound of Depth Estimation Error

We use Lemma 2 In order to compute the depth estimation error in a stereo system with non-uniform discretization.

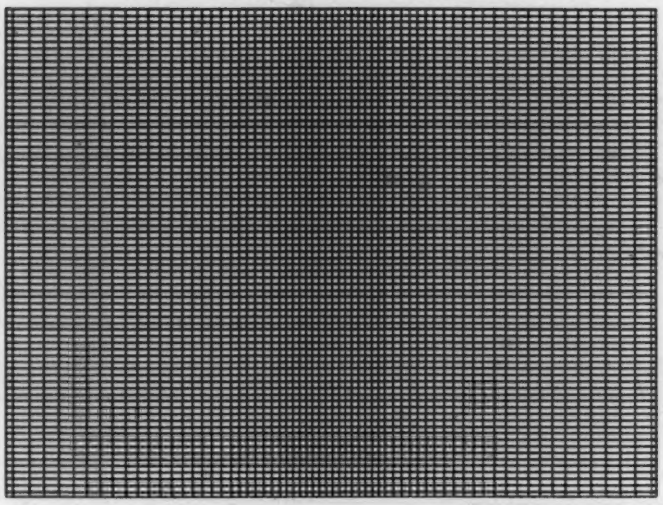


Figure 4.1: Exponential pixel distribution in the image plane for $\gamma = 0.03$

Lemma 2 *Suppose we have a single 3D point $P(X,Y,Z)$ which has projections x_r and x_l on the right and left cameras, respectively. Each camera has a non-uniform discretization; such a discretization is specified by a function $e_x(i)$ which defines the separation of pixels based on their location. The upper bound on the relative error in the depth estimation of P is then given from the equation below. For definition of parameters refer to Section 3.1 and List of Symbols.*

$$
\begin{aligned}
E_Z = \left| \frac{\hat{Z} - Z}{Z} \right| \leq\ & \frac{e_x(x_l)\sin\alpha_L}{2f}\,\frac{Z\cos\alpha_L - (X - dX)\sin\alpha_L}{Z} + \\
& \frac{e_x(x_r)\sin\alpha_R}{2f}\,\frac{Z\cos\alpha_R + X\sin\alpha_R}{Z} + \\
& \frac{e_x(x_l)\sin\alpha_L}{2fdX}\,\frac{X(Z\cos\alpha_L - (X - dX)\sin\alpha_L)}{Z} + \\
& \frac{e_x(x_r)\cos\alpha_R}{2fdX}(Z\cos\alpha_R + X\sin\alpha_R) + \\
& \frac{e_x(x_l)\cos\alpha_L}{2fdX}(Z\cos\alpha_L - (X - dX)\sin\alpha_L) - \\
& \frac{e_x(x_r)\sin\alpha_R}{2fdX}\,\frac{(X - dX)(Z\cos\alpha_R + X\sin\alpha_R)}{Z}
\end{aligned}
\tag{4.4}
$$

Proof: The proof of this lemma is similar to that of Theorem 2 in Chapter 3.

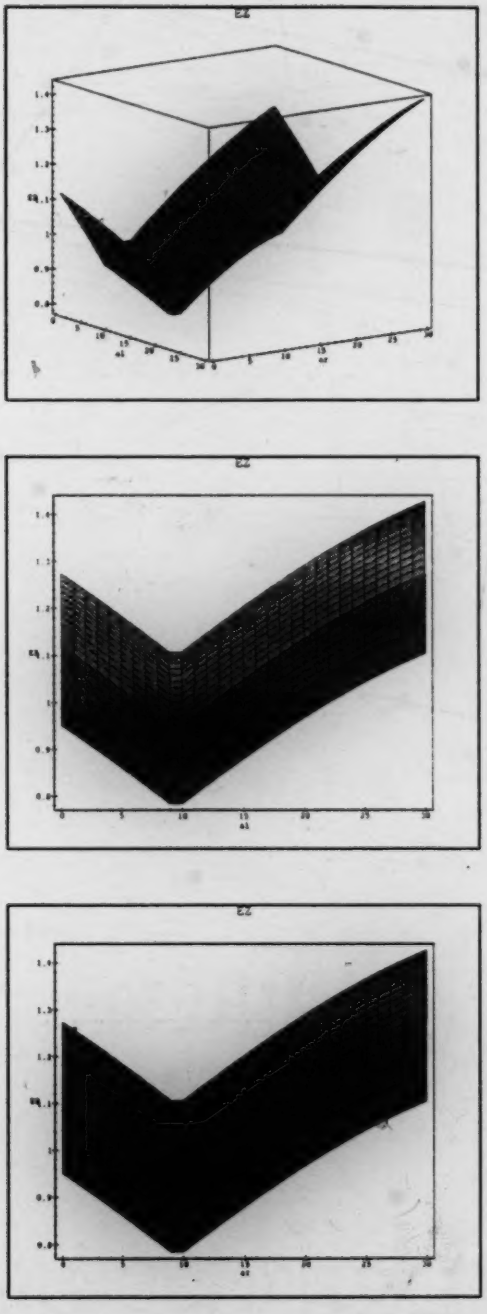


Figure 4.2: Upper bound of depth estimation error with non-uniform images for point $P(50, Y, 300)$

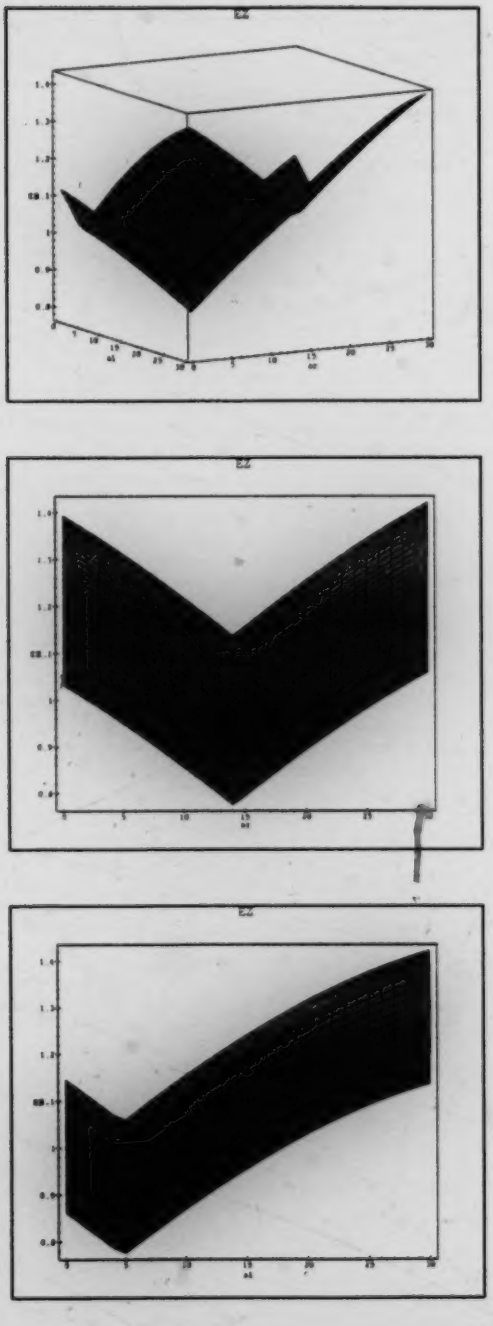


Figure 4.3: Upper bound of depth estimation error with non-uniform images for point $P(25, Y, 300)$

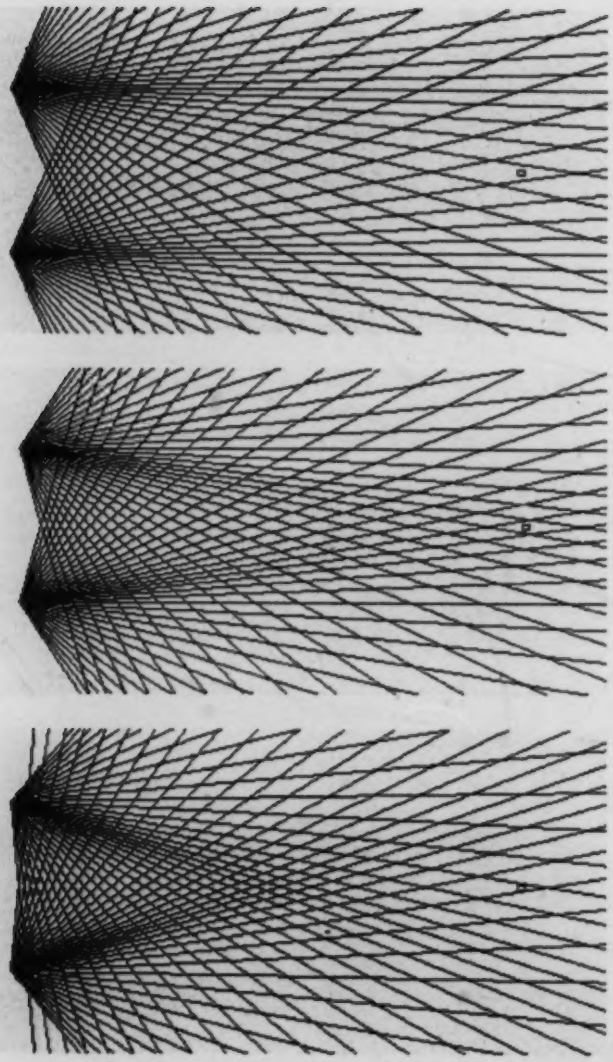


Figure 4.4: Variation of depth uncertainty with vergence using non-uniform image planes

Suppose in the system described in Chapter 3, we use cameras with an exponential pixel arrangement ($E_{min} = 0.025$, $\gamma = 0.01$) instead of cameras with uniform discretization. Using Lemma 2, the depth estimation error of point P located at X=50 mm and Z=300 mm is illustrated in Figure 4.2. The upper bound on the depth estimation error of P is *minimum* for the same vergence angle at which the error in the uniform resolution system was *maximum*; this is the vergence angle for which the projection of P lies on the center of the image plane of the cameras. Figure 4.4 illustrates the variation of depth uncertainty of a point P for three different vergence angles when cameras having non-uniform discretization. The plot of Figure 4.2 is graphically verified in terms of the size of diamonds in Figure 4.4.

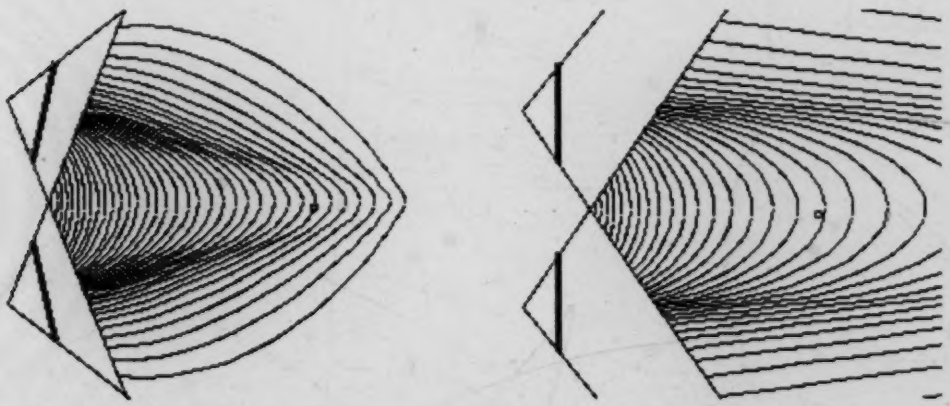


Figure 4.5: Isoresolution plots for stereo cameras with non-uniform discretization

Figure 4.5 gives isoresolution plots for stereo cameras with non-uniform exponential resolution when the point P is located symmetrically respect to two cameras. The density of the plots changes with vergence, in accordance with the theoretical results and the diamond plots.

We calculate the average depth estimation error for an object of interest in the scene using the same constraints on the X and Z dimensions of the object used in Chapter 3. The average depth estimation error for all the points lying on an object is calculated as:

$$\overline{E_Z} = \frac{\int_{Xmin}^{Xmax} \int_{Zmin}^{Zmax} E_Z dX dZ}{\int_{Xmin}^{Xmax} \int_{Zmin}^{Zmax} dX dZ} \tag{4.5}$$

where E_Z is obtained from Equation 4.4.

Suppose the object of interest (in a system with f = 10 mm, 40 mm × 40 mm imaging area, e_x = 0.025 mm and dX = 100 mm) is located in the region $280 < Z < 320$ and $35 < X < 65$. Figure 4.6 shows the average error for this object as a function of vergence angle. The minimum error occurs when the center of the object region is projected into the center of the image plane. The error of depth perception is obviously higher for peripherally visible targets.

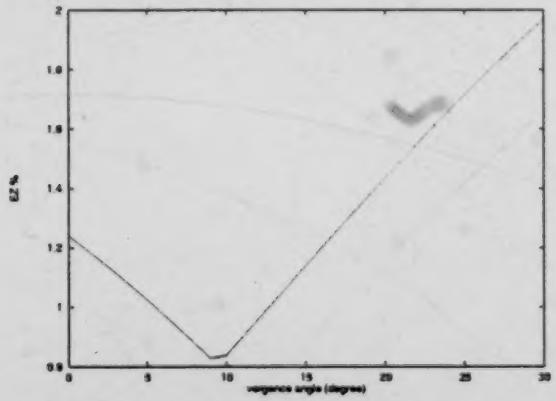


Figure 4.6: Average depth estimation error using non-uniform images

Figure 4.7 illustrates the variation of depth uncertainty when stereo cameras have non-symmetrical vergence angle with point P located non-symmetrically respect to the two cameras. The error in depth perception will be minimum when the pro-

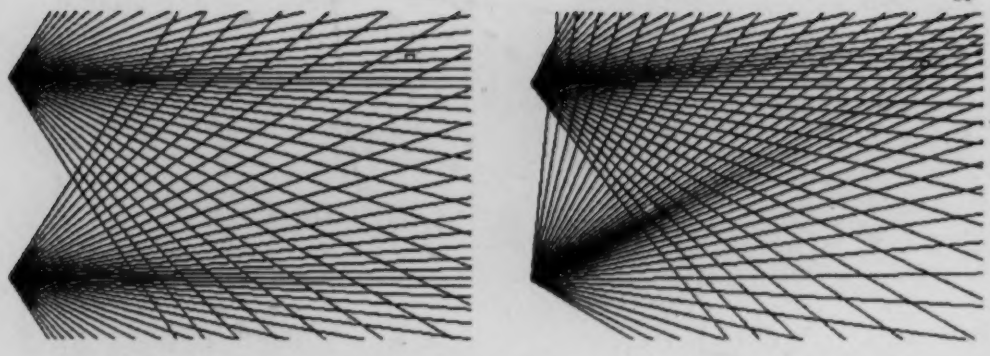


Figure 4.7: Variation of depth uncertainty for stereo cameras with non-uniform discretization and non-symmetrical vergence angles

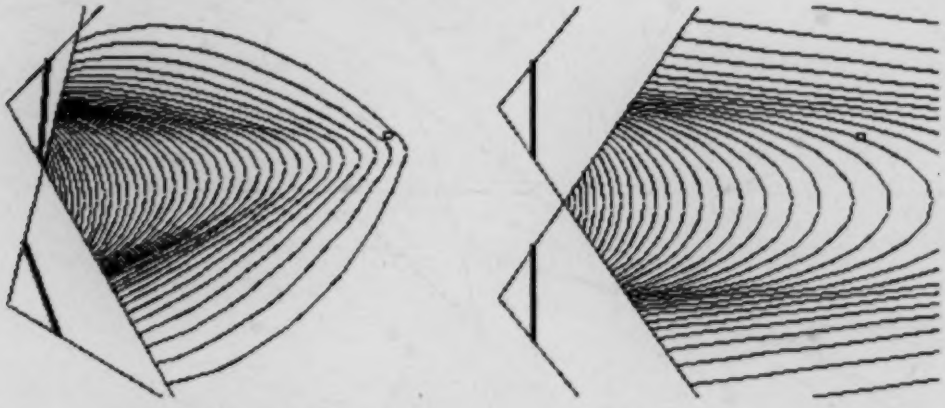


Figure 4.8: Isoresolution plots for stereo cameras with uniform discretization, and non-symmetrical vergence angles

jections of the point P in the scene are in the center of the image plane. The result is also verified by the isoresolution plots illustrated in Figure 4.8.

An issue that needs investigation is how the upper bound on error changes with changes in parameter γ in a non-uniform exponential system. In order to do so, we must first define the *focal angle* of a pixel. If each of the left and right boundaries of a pixel is connected by a line to the focal point of the camera, the two resulting lines create the focal angle of the pixel. These angles are illustrated in the diamond plots in this thesis. By studying the diamond plots from the uniform resolution system, we

observe that the behavior of depth perception error with vergence is directly related to the focal angles of the pixels. For the uniform resolution cameras, the focal angles of the pixels in the center of the image plane are larger than the focal angle of pixels in the periphery. This is why the depth perception error is larger when a vergence angle is selected such that the projection of the 3D point is in the center of the camera. For the non-uniform exponential resolution, there is no value for parameter γ that results in equal focal angles for all the pixels, and therefore the error plot versus vergence is never flat. However, the parameter γ should be chosen such that the focal angle of pixels increases monotonically, going from the center of an image plane towards the periphery. For small values of γ, although the pixel separation increases monotonically from the center toward the periphery, the focal angles may initially decrease and then increase — resulting in depth perception error changing in a non-monotonic fashion. Lemma 3 provides the minimum value of γ that results in a monotonic increase of focal angles from the center toward the periphery. The important point to note here is that the value of γ is derived from the geometry of the camera and not from the characteristics of the scene.

Lemma 3 *The minimum value of γ that ensures a monotonic increase in focal angles of pixels from the center of the image plane toward the periphery is:*

$$\gamma_{min} = \ln\left(\frac{2f^2}{f^2 - E_{min}^2} - 1\right) \tag{4.6}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Lemma 3 can be proved by finding the formula for focal angle as a function of the stereo imaging system parameters and studying it from elementary calculus.

—

Chapter 5

Optimal Non-uniform Discretization

In Chapter 4 we demonstrated that non-uniform discretization has a desirable effect on the depth estimation error with stereo camera's vergence movements. In this chapter, we study the optimal non-uniform discretization for stereo reconstruction. Once again, we assume that our cameras have uniform discretization in the Y direction and non-uniform discretization only along the X direction. Section 5.1 describes the error in estimation of X and Y coordinates an arbitrary 3D point in the scene. The errors in the X and Y coordinates are used to obtain the optimal non-uniform discretization. Section 5.2 and 5.3 present the optimal non-uniform exponential and linear discretization, respectively. Section 5.4 compares the resulting error from the optimal exponential case as an example with the error in a stereo system built using cameras with a uniform discretization.

5.1 Error in Estimation of X and Y Coordinates

In order to obtain optimal non-uniform discretization, we study the error in estimating the X and Y coordinates of an arbitrary point P as well as its depth(Z). Also, for the sake of simplicity, we restrict our study to the symmetrical vergence angles which can always be achieved by moving the stereo head toward the object rather than individual cameras separately. Lemma 4 provides equations for calculating the X and Y coordinates of the point P in Figure 3.2 from the coordinates of its projections in the right and left cameras and the vergence angle.

Lemma 4 *The X and Y coordinates of the point P are calculated from the following equations:*

$$X = Z\frac{(x_r \cos\alpha + f \sin\alpha)}{(f\cos\alpha - x_r \sin\alpha)} \tag{5.1}$$

$$Y = \frac{y_r(Z\cos\alpha + X\sin\alpha)}{f} \tag{5.2}$$

$$= \frac{y_l(Z\cos\alpha - (X - dX)\sin\alpha)}{f}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Lemma 4 is proved based on the method of Theorem 1, where a new coordinate system for each camera is defined, as illustrated in Figure 3.2. The origin of the new coordinate system is located in the focal point of the camera and its Z axis is perpendicular to the image plane. Let (X_R, Y_R, Z_R) and (X_L, Y_L, Z_L) be the coordinates of P in the new coordinate system that correspond to the right and left cameras, respectively. From equation 3.5, X can be obtained as a function of x_r and Z. Thus we have:

$$X = Z\frac{(x_r \cos\alpha + f \sin\alpha)}{(f\cos\alpha - x_r \sin\alpha)} \tag{5.3}$$

Finally, from trigonometry we have:

$$Y = \frac{Z_R y_r}{f} = \frac{Z_L y_l}{f} \tag{5.4}$$

By substituting for Z_R and Z_L in the above equation, equation 5.2 is obtained.

Because of the discretization error in approximating the location of projection of point P in the left and right cameras, there would be an error in estimating the coordinates of point P. In this study, we assume that the stereo cameras have non-uniform resolution in the X direction, and the resolution is uniform for both cameras in the Y direction. Theorem 3 demonstrates the relationship of the error in each of the coordinates with the pixel separation e_x and e_y. In this theorem, e_x is a function of location of projection and e_y is a constant.

Theorem 3 *An upper bound on the relative error in the X and Y coordinates of point $P(X,Y,Z)$ is given by the following equations:*

$$E_X = \left| \frac{\hat{X} - X}{X} \right| \leq E_Z + \frac{e_x(x_r)\cos\alpha(Z\cos\alpha + X\sin\alpha)}{2fX} + \frac{e_x(x_r)\sin\alpha(Z\cos\alpha + X\sin\alpha)}{2fZ} \tag{5.5}$$

$$E_Y = \left| \frac{\hat{Y} - Y}{Y} \right| \leq \frac{ZE_Z\cos\alpha + XE_X\sin\alpha}{Z\cos\alpha + X\sin\alpha} + \frac{e_y}{2y_r} + \frac{e_y(ZE_Z\cos\alpha + XE_X\sin\alpha)}{2y_r(Z\ cos\alpha + X\sin\alpha)} \tag{5.6}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: In order to obtain the maximum relative error in X, we use equation 5.1 from lemma 3. We have:

$$\hat{X} = \hat{Z}\frac{\hat{x}_r \cos\alpha + f\sin\alpha}{f\cos\alpha - \hat{x}_r \sin\alpha} \tag{5.7}$$

By substituting for $\hat{Z}$ and $\hat{x}_r$ in the above equation we have:

$$\begin{aligned}\hat{X} &= Z(1 \pm E_Z)[\frac{x_r \cos\alpha + f\sin\alpha \pm e_x(x_r)/2\cos\alpha}{f\cos\alpha - x_r\sin\alpha \pm e_x(x_r)/2\sin\alpha}] \\ &= Z[\frac{x_r\cos\alpha + f\sin\alpha}{f\cos\alpha - x_r\sin\alpha}][\frac{1 \pm e_x(x_r)/2\cos\alpha/(x_r\cos\alpha + f\sin\alpha)}{1 \pm e_x(x_r)/2\sin\alpha/(f\cos\alpha - x_r\sin\alpha)}]\end{aligned} \tag{5.8}$$

Therefore the upper bound of the relative error of X can be obtained as follows:

$$E_X = |\frac{\hat{X} - X}{X}| \leq (1 \pm E_Z)(1 \pm \frac{e_x(x_r)\cos\alpha}{2(x_r\cos\alpha + f\sin\alpha)})(1 \pm \frac{e_x(x_r)\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}) - 1$$

By substituting for x_r in the above equation, we have:

$$\begin{aligned}E_X^* = |\frac{\hat{X} - X}{X}| \leq (1 \pm E_Z)(1 \pm \frac{e_x(x_r)\cos\alpha(Z\cos\alpha + X\sin\alpha)}{2fX}) \\ (1 \pm \frac{e_x(x_r)\sin\alpha(Z\cos\alpha + X\sin\alpha)}{2fX}) - 1\end{aligned} \tag{5.9}$$

By expanding the right hand side of the above equation the upper bound on the relative error in X is obtained. Some of the terms are significantly smaller and can be ignored yielding equation 5.5.

In order to obtain the upper bound on the relative error in Y, we use equation 5.2. We use the equation that involves y_r, therefore we have:

$$\hat{Y} = \frac{\hat{y}_r(\hat{Z}\cos\alpha + \hat{X}\sin\alpha)}{f} \tag{5.10}$$

The projection of point P in each camera has at most $e_y/2$ discretization error. Since, the discretization in the Y direction is uniform, e_y is independent of the position of projection on the camera. We have:

$$\hat{y}_r = y_r \pm e_y/2 \qquad \hat{y}_l = y_l \pm e_y/2 \tag{5.11}$$

By substituting for $\hat{Z}$ and $\hat{y}_r$ in equation 5.10, we have:

$$\hat{Y} = \frac{Z(1 \pm E_Z)\cos\alpha + X(1 \pm E_X)\sin\alpha}{f}(y_r \pm e_y/2) \tag{5.12}$$

Thus the upper bound on the relative error in Y can be obtained as follows:

$$E_Y = |\frac{\hat{Y} - Y}{Y}| \leq \frac{ZE_Z\cos\alpha + XE_X\sin\alpha}{Z\cos\alpha + X\sin\alpha} + \frac{e_y}{2y_r} + \frac{e_y(ZE_Z\cos\alpha + XE_X\sin\alpha)}{2y_r(Z\ cos\alpha + X\sin\alpha)} \tag{5.13}$$

Now let us define the goal for obtaining optimal non-uniform discretization. The objective of the study is to find how to arrange pixels in the image plane to have a minimum upper bound on error of X, Y, and Z while resolution remains constant.

Let us assume that R denotes the resolution (the number of pixels) in a camera's image plane, which for simplicity, we consider to lie in a unit area. In the case of cameras with uniform resolution the following relationship between R, e_x and e_y exists:

$$(\frac{1}{e_x})(\frac{1}{e_y}) = R \tag{5.14}$$

If the resolution is non-uniform in the X direction, then the following relationship exists:

$$N(\frac{1}{e_y}) = R \tag{5.15}$$

where N is the number of pixels in the X direction in a unit length and we have:

$$1 = \sum_{i=0}^{N} e_x(i) \tag{5.16}$$

In order to obtain optimal vertical and horizontal pixel distribution in a non-uniform resolution stereo imaging system, the equations for E_X, E_Y and E_Z from

Theorem 2 and 3 must be considered. An obvious solution to minimize error in all directions is to reduce e_x and e_y together. However reducing e_x and e_y together leads to higher resolution, and that does not satisfy our objective; comparing the resulting error from two cameras with different resolution is not a fair comparison. To achieve minimum E_X and E_Z requires e_x in both cameras to be as small as possible. However when considering fixed resolution for the cameras, reducing e_x makes e_y larger and this is the worst possible solution for E_Y. Therefore in this study we turn our attention to a solution for minimization of E_Y. The solution will be optimal for E_Y; furthermore, E_X and E_Z will be less than what they would be if the cameras had simply followed industrial standards.

5.2 Optimal Exponential Discretization

In this section we study optimal non-uniform discretization modeled by equation 4.2; the pixel separation increases exponentially from the center of image plane toward the periphery. In order to calculate the upper bounds E_X, E_Y, and E_Z, the values of e_x and e_y are required. Theorem 4 provides us with the equations for e_x and e_y based on the resolution R, model parameters E_{min} and γ, and the coordinates of point $P(X, Y, Z)$. The value of e_x is different for the left and right cameras due to the difference in the location of the projections of point P, and non-uniform discretization in X direction. The value of e_y is independent of projection location due to uniform discretization.

Theorem 4 *The pixel separations e_x and e_y for projections of point P onto the left and right cameras with non-uniform exponential discretization modeled by equation*

4.2 are obtained from the following equations:

$$e_x(x_r) = \frac{E_{min}}{e^\gamma}[1 + \frac{f(X\cos\alpha - Z\sin\alpha)(e^\gamma - 1)}{(Z\cos\alpha + X\sin\alpha)E_{min}}] \tag{5.17}$$

$$e_x(x_l) = \frac{E_{min}}{e^\gamma}[1 + \frac{f((X - dX)\cos\alpha + Z\sin\alpha)(e^\gamma - 1)}{(Z\cos\alpha - (X - dX)\sin\alpha)E_{min}}] \tag{5.18}$$

$$e_y = \frac{1}{R}[\frac{1}{\gamma}\ln(1 + \frac{e^\gamma - 1}{E_{min}}) - 1] \tag{5.19}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Let us consider the right camera; we assume that there are m pixels between the center of the image plane and the projection of point P along the X direction. Therefore the following equation exists:

$$x_r = \sum_{i=0}^{m}(E_{min}e^{\gamma i}) \tag{5.20}$$

Computing the summation on the right hand side we have:

$$x_r = E_{min}\frac{(e^{\gamma(m+1)} - 1)}{e^\gamma - 1} \tag{5.21}$$

from which m can be obtained as:

$$m = \frac{1}{\gamma}\ln(1 + \frac{x_r(e^\gamma - 1)}{E_{min}}) - 1 \tag{5.22}$$

Substituting m from the above equation and x_r into equation 4.2 we have:

$$e_x(x_r) = E_{min}e^{\ln(1+\frac{x_r(e^\gamma-1)}{E_{min}})-\gamma} \tag{5.23}$$

$$= E_{min}e^{\ln(1+\frac{f(X\cos\alpha - Z\sin\alpha)(e^\gamma-1)}{(Z\cos\alpha+X\sin\alpha)E_{min}})-\gamma} \tag{5.24}$$

$$= \frac{E_{min}}{e^\gamma}[1 + \frac{f(X\cos\alpha - Z\sin\alpha)(e^\gamma - 1)}{(Z\cos\alpha + X\sin\alpha)E_{min}}] \tag{5.25}$$

Similarly for $e_x(x_l)$ we have:

$$e_x(x_l) = E_{min}e^{\ln(1+\frac{x_l(e^\gamma-1)}{E_{min}})-\gamma} \tag{5.26}$$

$$= E_{min}e^{\ln(1+\frac{f((X-dX)\cos\alpha+Z\sin\alpha)(e^\gamma-1)}{(Z\cos\alpha-(X-dX)\sin\alpha)E_{min}})-\gamma} \tag{5.27}$$

$$= \frac{E_{min}}{e^\gamma}[1 + \frac{f((X - dX)\cos\alpha + Z\sin\alpha)(e^\gamma - 1)}{(Z\cos\alpha - (X - dX)\sin\alpha)E_{min}}] \tag{5.28}$$

The number of pixels in the x direction of image plane in unit length (N) is obtained as follows:

$$N = \frac{1}{\gamma}\ln(1 + \frac{1 \times (e^{\gamma} - 1)}{E_{min}}) - 1 \tag{5.29}$$

Subsequently we have:

$$e_y = \frac{1}{R}[\frac{1}{\gamma}\ln(1 + \frac{1 \times (e^{\gamma} - 1)}{E_{min}}) - 1] \tag{5.30}$$

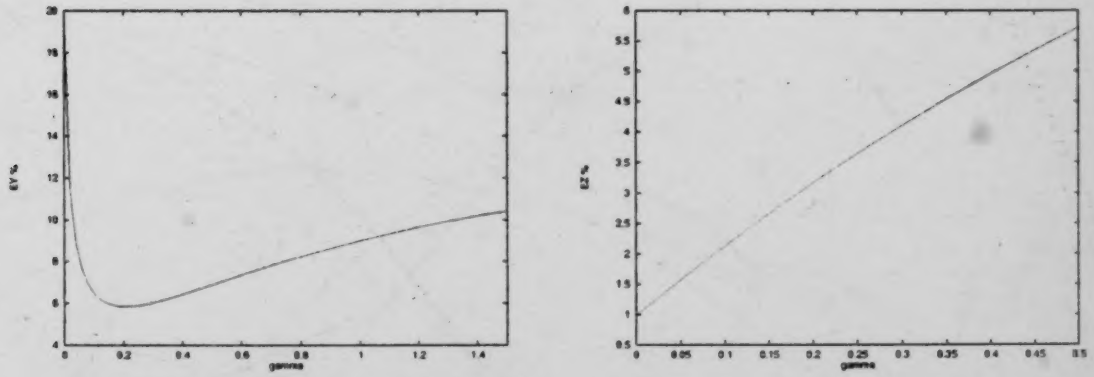


Figure 5.1: The error EY (left) and EZ (right) for an object using stereo cameras with non-uniform exponential discretization.

By substituting equations 5.17, 5.18, 5.19 into equation 5.6, the value of E_Y for the point $P(X, Y, Z)$ is obtained. The average error for an object in the scene located within $Zmin < Z < Zmax$ and $Xmin < X < Xmax$ is obtained from the following equation:

$$\overline{E_Y} = \frac{\int_{Xmin}^{Xmax}\int_{Zmin}^{Zmax} E_Y dX dZ}{\int_{Xmin}^{Xmax}\int_{Zmin}^{Zmax} dX dZ} \tag{5.31}$$

Let us assume a stereo system where f is 20 mm, dx is 100 mm, E_{min} is 0.05 mm and resolution R is 1000, the average error $\overline{E_Y}$ for the object located within $350 < Z < 400$ and $25 < X < 75$ boundaries is plotted in Figure 5.1. The value of γ corresponding to minimum $\overline{E_Y}$ is derived from this plot.

5.3 Optimal Linear Discretization

In this section we study optimal non-uniform discretization modeled by equation 4.3; the pixel separation increases linearly from the center of the image plane toward the periphery. Theorem 5 provides us with the equations for e_x and e_y for the projection of a point $P(X, Y, Z)$ in both cameras, based on the resolution R, model parameters E_{min} and ν, and the coordinates of point P.

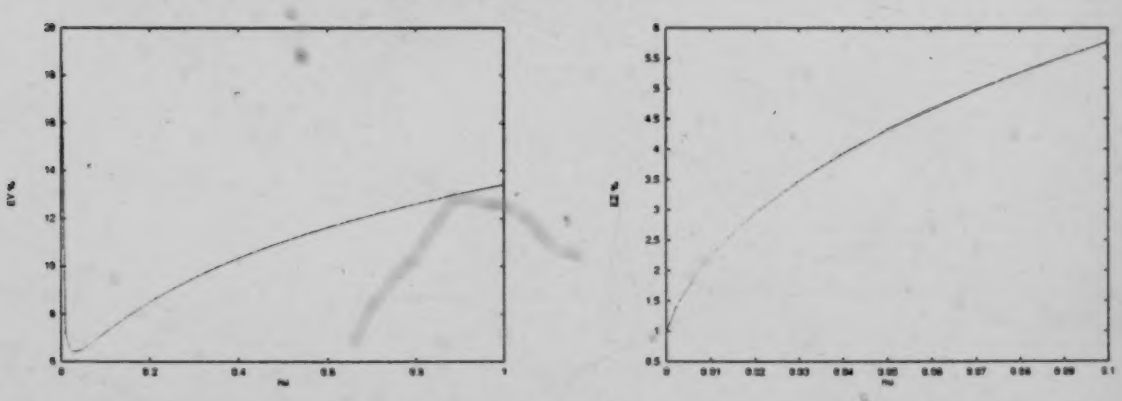


Figure 5.2: The error EY (left) and EZ (right) for an object using stereo cameras with non-uniform linear discretization

Theorem 5 *The pixel separation e_x and e_y for the projection of point P onto the left and right cameras with non-uniform discretization modeled by equation 4.3 are obtained from the following equations:*

$$e_x(x_r) = E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - f\frac{(X\cos\alpha - Z\sin\alpha)}{(Z\cos\alpha + X\sin\alpha)})} - (\nu + 2E_{min})}{2} \quad (5.32)$$

$$e_x(x_l) = E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - f\frac{((X-dX)\cos\alpha + Z\sin\alpha)}{(Z\cos\alpha - (X-dX)\sin\alpha)})} - (\nu + 2E_{min})}{2} \quad (5.33)$$

$$e_y = \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - 1)} - (\nu + 2E_{min})}{2\nu R} \quad (5.34)$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Let us first consider the right camera. We assume there are m pixels between the center of the image plane and the projection of point P along the X direction. We have:

$$x_r = \sum_{i=0}^{m} (E_{min} + \nu i) \tag{5.35}$$

Computing the summation on the right hand side we have:

$$x_r = (m+1)E_{min} + \frac{m(m+1)\nu}{2} \tag{5.36}$$

which can be expressed as:

$$\nu m^2 + (\nu + 2E_{min})m + 2(E_{min} - x_r) = 0 \tag{5.37}$$

The positive root of this equation is:

$$m = \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - x_r)} - (\nu + 2E_{min})}{2\nu} \tag{5.38}$$

Substituting m from the above equation and x_r into equation 4.3, we have:

$$e_x(x_r) = E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - x_r)} - (\nu + 2E_{min})}{2} \tag{5.39}$$

$$= E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - f\frac{(X\cos\alpha - Z\sin\alpha)}{(Z\cos\alpha + X\sin\alpha)})} - (\nu + 2E_{min})}{2} \tag{5.40}$$

Similarly for $e_x(x_l)$:

$$e_x(x_l) = E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - x_l)} - (\nu + 2E_{min})}{2} \tag{5.41}$$

$$= E_{min} + \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - f\frac{((X - dX)\cos\alpha + Z\sin\alpha)}{(Z\cos\alpha - (X - dX)\sin\alpha)})} - (\nu + 2E_{min})}{2} \tag{5.42}$$

Let us denote the number of pixels in the unit length in the x direction as N; it is obtained from the following equation:

$$N = \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - 1)} - (\nu + 2E_{min})}{2\nu} \tag{5.43}$$

The following relationship exists between N and the resolution R (which is assumed to be constant).

$$R = N\frac{1}{e_y} \tag{5.44}$$

e_y can be obtained from the equations 5.44 and 5.43 as follows:

$$e_y = \frac{\sqrt{(\nu + 2E_{min})^2 - 8\nu(E_{min} - 1)} - (\nu + 2E_{min})}{2\nu R} \tag{5.45}$$

By substituting equations 5.32, 5.33, 5.34 into equation 5.6, the value of E_Y for a point $P(X, Y, Z)$ is obtained. Suppose that f is 20 mm, dx is 100 mm, E_{min} is 0.05 mm and the resolution R is 1000. Figure 5.2 (left) illustrates the graph of E_Y versus ν for a small object located within the $350 < Z < 400$ and $25 < X < 75$ boundaries when two cameras are tilted toward the object.

By substituting equations 5.32, 5.33, 5.34 into equation 4.4, we obtain the E_Z value for a point $P(X, Y, Z)$. For the example in the previous section, the variation of $\overline{E_Z}$ with ν is illustrated in Figure 5.2 (right) $\overline{E_Z}$ increases directly with the increase in ν, as expected.

5.4 Comparison of Uniform and Non-Uniform Discretization

In the previous sections, we found the optimal non-uniform discretization in the X direction. The significance of these results is only understood when the error is compared with the uniform resolution case. For non-uniform exponential discretization, the optimal value of EY is obtained for $\gamma = 0.2$. For this value of γ, $EY = 5.85\%$

and $EZ = 3.17\%$. Figure 5.3 illustrates the upper bound on average error EY as a function of pixel separation e_x for the example system in this chapter when both stereo cameras have uniform resolution. The error EY in the best case is above 8.5%, which is well above the values obtained by non-uniform resolution case. This best value of EY is obtained for $e_x = 0.22$. Figure 5.4 illustrates the error EZ for this value of e_x as a function of vergence angle. The value of EZ is maximum for the vergence angle where both cameras are tilted toward the object. For this vergence angle, its value $EZ = 3.82$ is still more than the values obtained by the non-uniform resolution. In other words the optimal non-uniform discretization not only improves EY over the uniform resolution, but also EZ is improved.

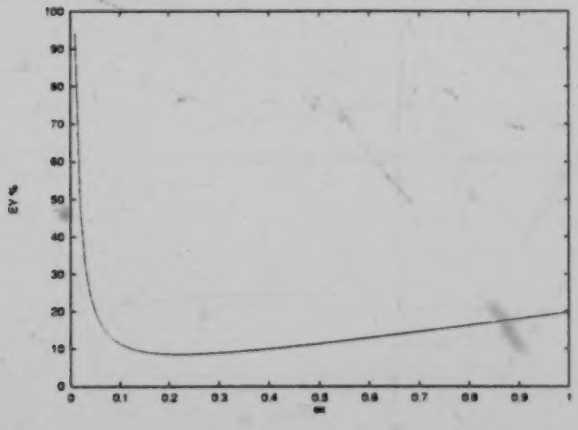


Figure 5.3: EY for using stereo cameras with uniform discretization.

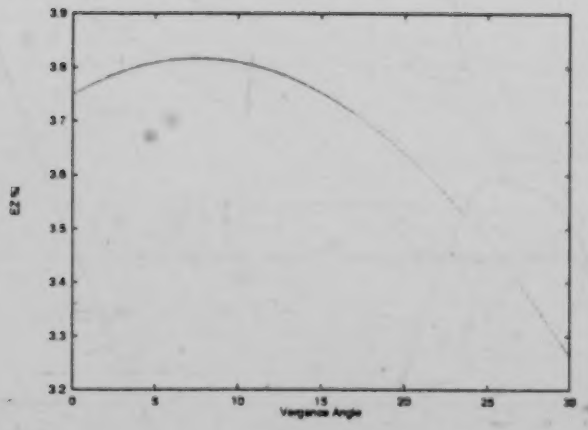


Figure 5.4: EZ for using stereo cameras with uniform discretization for $e_x = 0.22$.

Chapter 6

Stereo Cameras with Tilt Movements

In this chapter we examine a stereo imaging system with a less restricted configuration, where cameras have independent tilt as well as pan movements similar to the human eye. The object of interest can be located anywhere in 3D space with respect to the cameras and not necessarily symmetrically located before the two cameras.

Section 6.1 illustrates the configuration of the stereo system which is studied in this chapter. The depth of an arbitrary 3D point in the scene will be calculated based on the stereo imaging system parameters. In Section 6.2, we study the error in depth estimation. Section 6.3 discusses the epipolar lines computation in this stereo imaging system which is used for stereo matching.

6.1 Depth of A 3D Point

Figure 6.1 illustrates the configuration of the stereo imaging system with cameras having independent pan and tilt movements. The pan angle, which is the angle of

rotation of the camera's coordinate system along the Y axis, is denoted by β and θ for the left and right cameras respectively. The tilt angle which is the angle of rotation of the camera's coordinate system along the X axis, is denoted by α and γ for the left and right cameras respectively.

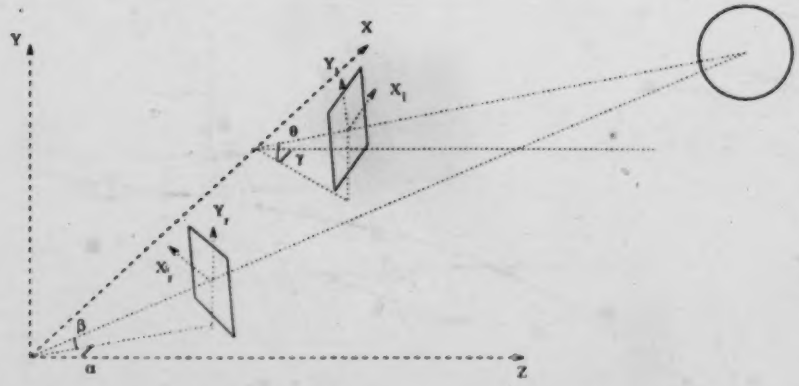


Figure 6.1: Stereo imaging system with tilt movements

Theorem 6 provides the equation for depth of a point P in Figure 6.1 based on the physical geometry of the stereo imaging system and the projection of point P in the left and right cameras.

Theorem 6 *The depth of a 3D point $P(X, Y, Z)$ in the stereo imaging system of Figure 6.1 can be obtained from the following equation:*

$$Z = \frac{Z_R Y_L dX}{Y_L X_R - (X_L - dX) Y_R} \tag{6.1}$$

in which:

$$Z_R = -y_r \sin\beta + \cos\beta(-x_r \sin\alpha + f\cos\alpha) \tag{6.2}$$

$$Z_L = -y_l \sin\theta + \cos\theta(-x_l \sin\gamma + f\cos\gamma)$$

$$X_R = x_r \cos\alpha + f\sin\alpha$$

$$X_L = x_l \cos\gamma + f\sin\gamma + dX$$

$$Y_R = y_r \cos\beta + \sin\beta(-x_r \sin\alpha + f\cos\alpha)$$

$$Y_L = y_l \cos\theta + \sin\theta(-x_l \sin\gamma + f\cos\gamma) \tag{6.3}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: In order to find the depth (Z) of point $P(X, Y, Z)$ from the coordinates of its projections in the left and right cameras, we use the matrix form of the rotation, translation and projection transformations. The matrix form of the transformations is easily accomplished by using a homogeneous coordinate system. The homogeneous coordinates of a point with coordinates (X, Y, Z) are defined by (kX, kY, kZ, k) where k is an arbitrary nonzero constant.

Let us consider a point $P_R(x_r, y_r)$ in the right camera's image plane. Assuming that the right camera pan and tilt angle is zero, point P_R has 3D coordinates (x_r, y_r, f) in the world 3D coordinate system with its origin in the right camera's focal point. Using the matrix form of the rotation and homogeneous coordinates of point P_R, if the right camera has nonzero pan and tilt angles, then the 3D coordinates of point P_R are obtained as follows:

$$C_{hR} = R_\beta R_\alpha W_{hR} \tag{6.4}$$

where W_{hR} denotes the homogeneous coordinate of point P_R, and C_{hR} represents the resulting point homogeneous coordinate after transformation. R_α is the rotation matrix along the Y axis, and R_β is the rotation matrix along the X axis as follows:

$$R_\alpha = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\alpha & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad R_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta & \sin\beta & 0 \\ 0 & -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.5}$$

Similarly, we have the following equation for the point P_L in the left camera's image plane:

$$C_{hL} = R_\theta R_\gamma C W_{hL} \tag{6.6}$$

where, similar to P_R, W_{hL} denotes the homogeneous coordinate of point P_L, and C_{hL} represents the resulting point homogeneous coordinate after transformation. C is the displacement matrix, R_γ is the rotation matrix along the Y axis, and R_θ is the rotation matrix along the X axis as follows:

$$C = \begin{pmatrix} 1 & 0 & 0 & dX \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.7}$$

$$R_\gamma = \begin{pmatrix} \cos\gamma & 0 & \sin\gamma & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\gamma & 0 & \cos\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad R_\theta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.8}$$

Let us assume point $P(X, Y, Z)$ has projections $P_R(X_R, Y_R, Z_R)$ and $P_L(X_L, Y_L, Z_L)$ in the right and left cameras, respectively. Equation 6.10 describes all the 3D lines passing from P_R and the origin (which happens to be the focal point of the right camera) in parametric form:

$$X = X_R t \tag{6.9}$$

$$Y = Y_R t$$

$$Z = Z_R t$$

Similarly, Equation 6.11 demonstrates all the 3D lines passing from P_L and the focal point of the left camera, which is located at $(dX, 0, 0)$ in parametric form:

$$X = dX + (X_L - dX)t' \tag{6.10}$$

$$Y = Y_L t'$$

$$Z = Z_L t'$$

The point P projection line into the right camera passing from point P_R and P, and the projection line passing from point P_L and P have point P in common. Therefore point P should satisfy both equations, and we have:

$$t = \frac{Y_L dX}{Y_L X_R - (X_L - dX)Y_R} \tag{6.11}$$

$$t' = \frac{Y_R dX}{Y_L X_R - (X_L - dX)Y_R} \tag{6.12}$$

By substituting for the t in equation 6.10, we have:

$$Z = \frac{Z_R Y_L dX}{Y_L X_R - (X_L - dX)Y_R} \tag{6.13}$$

In order to obtain the relationship between the 3D coordinates of P_R and P_L with the coordinates of the projection on the image plane (e.g. X_R, Y_R, Z_R and x_r, y_r),

we use the matrix form of the projection transformation T_P as follows:

$$T_P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \tag{6.14}$$

Let us denote the homogeneous coordinate of point P with vector W_h and the resulting point homogeneous coordinate with vector C_h. Equation 6.15 provides the projection of point P in the projection plane defined by the equation $Z = f$, which coincides with our right camera's image plane with zero pan and tilt angles.

$$C_h = T_P W_h \tag{6.15}$$

The projection of point P in the cameras with nonzero α pan and β tilt angles can be calculated by first rotating point P by $-\alpha$ and $-\beta$ along the Y and X axes, respectively, and then finding the projection using the equation 6.15. This leads to the following equations for the right and left cameras:

$$C_{hR} = T_P R_{-\beta} R_{-\alpha} W_h \tag{6.16}$$

$$C_{hL} = T_P R_{-\theta} R_{-\gamma} C W_h \tag{6.17}$$

The projection point P_R with (x_r, y_r) in the right camera's image plane has coordinates (x_r, y_r, f) in the 3D world coordinate system, and similarly the projection point P_L with coordinates (x_l, y_l) in the left camera's image plane has coordinates (x_l, y_l, f) in the 3D world coordinate system. Considering these facts and expanding the above equations, the equation 6.1 for depth (Z coordinate) of point P is obtained.

6.2 Depth Estimation Error

Because of discretization error in approximating of the location of projection of point P in the left and right cameras, there is an error estimating the coordinates of point P. In this study, we assume that the stereo cameras have uniform discretization in the x and y direction, and each camera can have independent pan and tilt movements. Theorem 7 demonstrates the relationship of the error in the depth coordinate Z with pixel separation e.

Theorem 7 *The upper bound on relative error in depth of a 3D point $P(X,Y,Z)$ is obtained from the following equation:*

$$E_Z = \left| \frac{\hat{Z} - Z}{Z} \right| \leq \left| \frac{\Delta Z_R}{Z} \right| + \left| \frac{\Delta Y_L}{Y_L} + \frac{X_R \Delta Y_L + Y_L \Delta X_R - Y_R \Delta X_L - (X_L - dX)\Delta Y_R}{Y_L X_R - (X_L - dX) Y_R} \right| \tag{6.18}$$

in which equations 6.3 hold and we have:

$$\Delta Z_R = \pm e/2(\sin\beta + \cos\beta \sin\alpha) \tag{6.19}$$

$$\Delta Z_L = \pm e/2(\sin\theta + \cos\theta \sin\gamma)$$

$$\Delta X_R = \pm e/2 \cos\alpha$$

$$\Delta X_L = \pm e/2 \cos\gamma$$

$$\Delta Y_R = \pm e/2(\cos\beta + \sin\beta \sin\alpha)$$

$$\Delta Y_L = \pm e/2(\cos\theta + \sin\theta \sin\gamma)$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: In order to prove this theorem we use the following equation:

$$Z = Z_R t \tag{6.20}$$

in which t is obtained from equation 6.11. Due to discretization error, we have:

$$\hat{Z} = \hat{Z_R}\hat{t} \tag{6.21}$$

By expanding the right hand side, we have:

$$\hat{Z} = Z_R t(1 \pm \frac{\Delta t}{t})(1 \pm \frac{\Delta Z_R}{Z_R}) \tag{6.22}$$

By multiplying two terms together and ignoring the second order error terms we have:

$$E_Z = | \frac{\hat{Z} - Z}{Z} | \leq | \frac{\Delta Z_R}{Z} | + | \frac{\Delta t}{t} | \tag{6.23}$$

In this equation, ΔZ_R and ΔZ_L are unknowns. First we find the error in coordinates of the projection of the point $P(X, Y, Z)$ in the right and left camera $P_R(X_R, Y_R, Z_R)$ and $P_L(X_L, Y_L, Z_L)$ due to discretization. For the Z_R coordinate, we have:

$$Z_R = -y_r \sin\beta + \cos\beta(-x_r \sin\alpha + f\cos\alpha) \tag{6.24}$$

Due to discretization we have:

$$\hat{x_r} = x_r \pm e/2 \tag{6.25}$$

$$\hat{y_r} = y_r \pm e/2 \tag{6.26}$$

Since the cameras have uniform resolution, e is constant and we only need to consider the case where pixel separation is the same in both directions. We have:

$$\hat{Z_R} = -\sin\beta(y_r \pm e/2) + \cos\beta(-(x_r \pm e/2)\sin\alpha + f\cos\alpha) \tag{6.27}$$

Therefore we have:

$$\Delta Z_R = \pm e/2(\sin\beta + \cos\beta\sin\alpha)) \tag{6.28}$$

Similarly, for the rest of the coordinates we have:

$$\Delta Z_L = \pm e/2(\sin\theta + \cos\theta \sin\gamma) \tag{6.29}$$

$$\Delta X_R = \pm e/2 \cos\alpha$$

$$\Delta X_L = \pm e/2 \cos\gamma$$

$$\Delta Y_R = \pm e/2(\cos\beta + \sin\beta \sin\alpha)$$

$$\Delta Y_L = \pm e/2(\cos\theta + \sin\theta \sin\gamma)$$

The other unknown in the equation 6.23 is the Δt, which we calculate here:

$$t = \frac{Y_L dX}{Y_L X_R - (X_L - dX)Y_R} \tag{6.30}$$

Then we have:

$$\hat{t} = \frac{(Y_L + \Delta Y_L)dX}{(Y_L + \Delta Y_L)(X_R + \Delta X_R) - (X_L + \Delta X_L - dX)Y_R} \tag{6.31}$$

By expanding the right hand side of the above equation and ignoring the second order error terms and assuming error in denominator is small we have:

$$\Delta t \simeq t(\frac{\Delta Y_L}{Y_L} + \frac{X_R \Delta Y_L + Y_L \Delta X_R - Y_R \Delta X_L - (X_L - dX)\Delta Y_R}{DEN(t)}) \tag{6.32}$$

By substituting the above equation in equation 6.23, the equation 6.18 is obtained.

Theorem 7 presents an upper bound on the depth estimation error for a single point. We are interested in the average error for an object. The average depth estimation error for an object within the boundaries of $X_{min} < X < X_{max}, Y_{min} < Y < Y_{max}, Z_{min} < Z < Z_{max}$ is obtained as follows:

$$\bar{E} = \frac{\int_{X_{min}}^{X_{max}} \int_{Y_{min}}^{Y_{max}} \int_{Z_{min}}^{Z_{max}} E_Z dx dy dz}{\int_{X_{min}}^{X_{max}} \int_{Y_{min}}^{Y_{max}} \int_{Z_{min}}^{Z_{max}} dx dy dz} \tag{6.33}$$

Now let us consider a stereo imaging system consisting of two cameras which are separated by 100 mm (dX=100). Two cameras are identical with focal length of 10 mm, and the distance between centers of two adjacent pixels is 0.05 mm. Figure 6.2 illustrates the upper bound on the average relative error in depth estimation of an object situated within the following boundaries: $25 < X < 75$, $275 < Z < 325$, $25 < Y < 75$. This figure plots the error when the two cameras have a fixed 10 degree tilt angle toward the object, and the pan angles of both cameras varies towards the object. Figure 6.2 clearly illustrates that depth estimation error is maximum when the cameras are panned and tilted toward the object in the scene.

6.3 Epipolar Lines

In this thesis we are not directly concerned with the matching process in stereo images. However, most of the matching algorithms in literature are restricted to a simple parallel stereo imaging system. A stereo imaging system which each camera can have independent pan and tilt movements has not been fully studied. Theorem 8 derives the epipolar line equation in the left image for an arbitrary point in the right image — the match for the selected point resides on the derived epipolar line.

Theorem 8 *For the 3D point $P(X, Y, Z)$ with projection point $P_R(x_r, y_r)$ in the right image, the following equation is the epipolar line equation in the left camera's image plane corresponding to point P_R:*

$$\begin{aligned} & x_l(\sin\gamma\cos\gamma - C_Y\sin\theta\sin\gamma) + \\ & y_l(C_Y\cos\theta + f\sin^2\gamma - f) + \\ & C_Y f\sin\theta\cos\gamma = 0 \end{aligned} \tag{6.34}$$

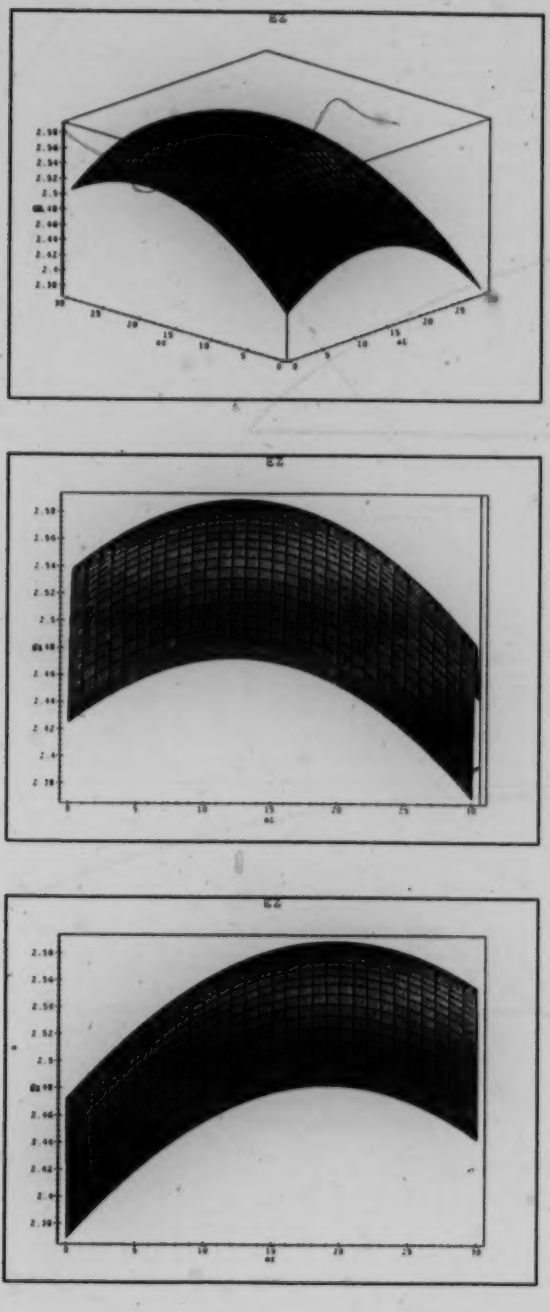


Figure 6.2: Upper bound of depth estimation error of location of a 3D object in stereo imaging

in which

$$C_Y = \cos\gamma\sin\theta - \cos\gamma\cos\theta\frac{Z_R}{Y_R} \tag{6.35}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Figure 6.3 illustrates the epipolar lines in our stereo configuration. Epipolar lines are derived from the intersection of the epipolar plane, which is the plane passing through the focal points of the right and left cameras and the 3D point P, and the image plane of the cameras.

Point $P_R(x_r, y_r)$ and the projection of point $P(X, Y, Z)$ have the following coordinates in the 3D coordinate system:

$$X_R = x_r\cos(\alpha) + f\sin(\alpha) \tag{6.36}$$

$$Y_R = y_r\cos(\beta) + \sin(\beta)(-x_r\sin(\alpha) + f\cos(\alpha))$$

$$Z_R = -y_r\sin(\beta) + \cos(\beta)(-x_r\sin(\alpha) + f\cos(\alpha))$$

The epipolar plane equation which passes focal points of left and right cameras and P_R is defined by the following equation:

$$dx(Z_R Y + Y_R Z) = 0 \tag{6.37}$$

The equation for the left camera's image plane is obtained by calculating the normal vector of the plane considering its pan and tilt movements, and selecting an arbitrary point on the plane. We have:

$$\begin{aligned} \sin(\gamma)(X - f\sin(\gamma) - dx) &+ \\ \cos(\gamma)\sin(\theta)(Y - f\cos(\gamma)\sin(\theta)) &+ \\ \cos(\gamma)\cos(\theta)(Z - f\cos(\gamma)\cos(\theta)) &= 0 \end{aligned} \tag{6.38}$$

The epipolar line equation is obtained from the intersection of the two planes described by equations 6.37 and 6.38 as follows:

$$X\sin(\gamma) - f - dx\sin(\gamma) + (\cos(\gamma)\sin(\theta) - \cos(\gamma)\cos(\theta)Z_R/Y_R)Y \;=\; 0 \quad (6.39)$$

in which X and Y are the 3D coordinates of projection points onto the left image plane. By substituting equation 6.3 into the above equation, the epipolar line equation 6.35 is obtained.

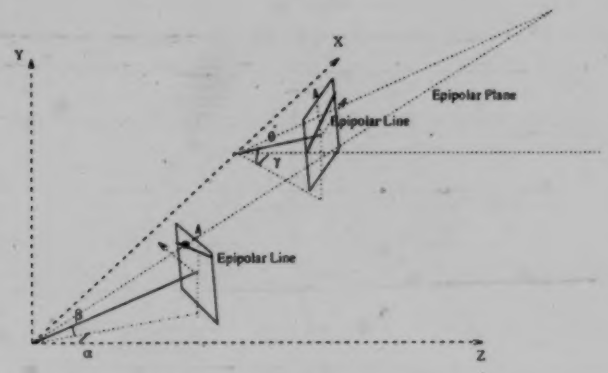


Figure 6.3: Epipolar lines in a stereo system with vergence

Chapter 7

Spatially-Varying Stereo Cameras with Tilt Movements

In this chapter we study the idea of non-uniform discretization presented in Chapter 4, for the stereo system described in Chapter 6. However, unlike Chapter 4, the cameras will have non-uniform discretization in both the horizontal and vertical directions.

Section 7.1 describes the upper bound of the error in the stereo system of Chapter 6 built using cameras with non-uniform exponential discretization. Section 7.2 discusses the optimal non-uniform discretization in the stereo imaging system of Chapter 6.

7.1 Upper Bound of Error in Z

Theorem 7 in Chapter 6 provides the upper bound for the depth estimation error when two cameras have uniform resolution. For the cameras with non-uniform resolution, Theorem 7 is still valid based on the values for $\Delta X_R, \Delta X_L, \Delta Y_R, \Delta Y_L, \Delta Z_R, \Delta Z_L$ provided by the following lemma:

Lemma 5 *The absolute error for the 3D coordinates of projection of point P in the left and right cameras is obtained from the following equations:*

$$\Delta Z_R = \pm 1/2(e(y_r)\sin\beta + e(x_r)\cos\beta\sin\alpha) \tag{7.1}$$

$$\Delta Z_L = \pm 1/2(e(y_l)\sin\theta + e(x_l)\cos\theta\sin\gamma)$$

$$\Delta X_R = \pm 1/2e(x_r)\cos\alpha$$

$$\Delta X_L = \pm 1/2e(x_l)\cos\gamma$$

$$\Delta Y_R = \pm 1/2(e(y_r)\cos\beta - e(x_r)\sin\beta\sin\alpha)$$

$$\Delta Y_L = \pm 1/2(e(y_l)\cos\theta - e(x_l)\sin\theta\sin\gamma)$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: Proof can be easily accomplished using equation 6.3 and expanding the right hand side using the error in the coordinates of projections. The error depends on the value of a coordinate rather than being a fixed value.

—

In this study we consider the exponential non-uniform discretization model described by the following equation:

$$e(i) = E_{min}e^{\nu i} \tag{7.2}$$

where E_{min} is the pixel separation at the center of the camera, ν is a constant coefficient that controls the rate of increase of the pixel separation, and i is the manhattan distance of the pixel along X or Y directions from the center pixel.

Now let us consider the stereo imaging system with identical setup described in the Chapter 6. We replace the cameras with ones that have non-uniform discretization described by equation 7.2 ($E_{min} = 0.01$ and $\nu = 0.01$). Note that these numbers are

selected so that the overall resolution of the cameras with non-uniform resolution in unit area is the same as the one with uniform resolution in the Chapter 6. This is done to make the comparison of uniform and non-uniform resolution a fair one. Figure 7.1 illustrates the variation of upper bound of depth estimation error for the object of interest in the scene. Figure 7.1 illustrates that the error will be minimum when the cameras are panned and tilted toward the object.

The significance of the result illustrated in Figure 7.1 is only perceived by comparison to the results illustrated in Figure 6.2. The upper bound in non-uniform resolution case when cameras are panned and tilted toward the object is far less than the uniform resolution case.

7.2 Optimal Non-Uniform Discretization

In section 7.1, we proved that non-uniform discretization similar to the human eye provides far better results in depth estimation error compared to cameras with uniform discretization. In this section we study optimal non-uniform discretization provided that we are restricted to the exponential non-uniform discretization model described in equation 7.2.

In the stereo imaging system described in the Chapter 6, we analyze average error in depth estimation for the object in the scene with the changes in ν, where the cameras are panned and tilted toward the object. Again, in order to have a fair comparison, with each change in the value of ν, the value of E_{min} is changed to keep the overall resolution at a fixed value. Figure 7.2 illustrates the change in average error versus ν. Figure 7.3 illustrates the change of E_{min} versus ν. For each object in the scene there is an optimal non-uniform discretization which leads to minimum

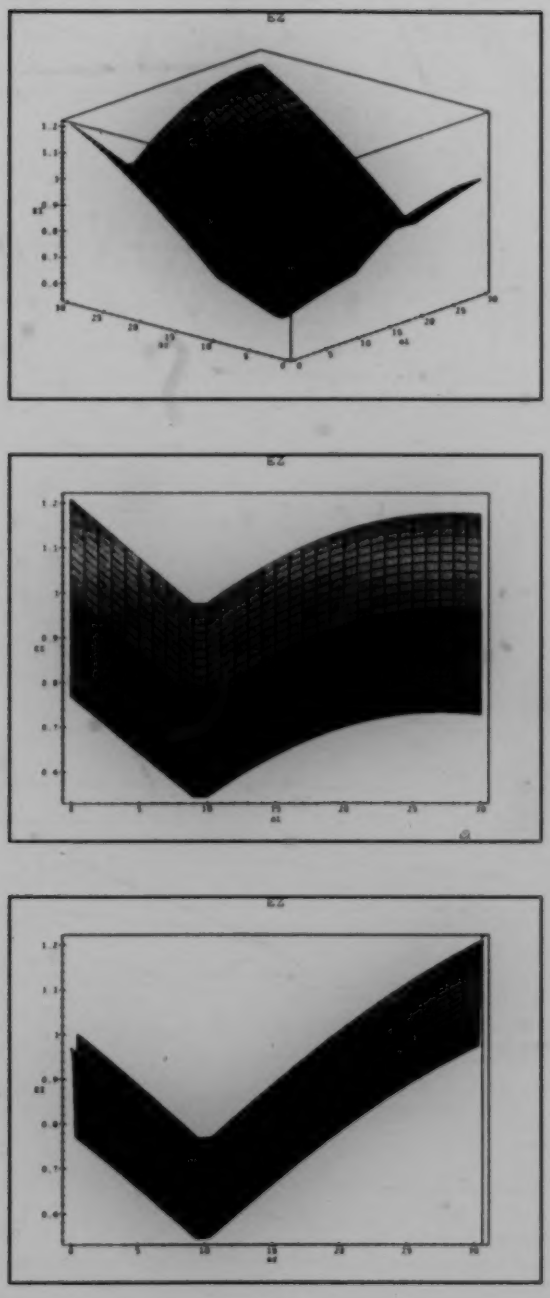


Figure 7.1: Upper bound on depth estimation error in the location of a 3D object in stereo imaging system with non-uniform pixel distribution

average depth estimation error.

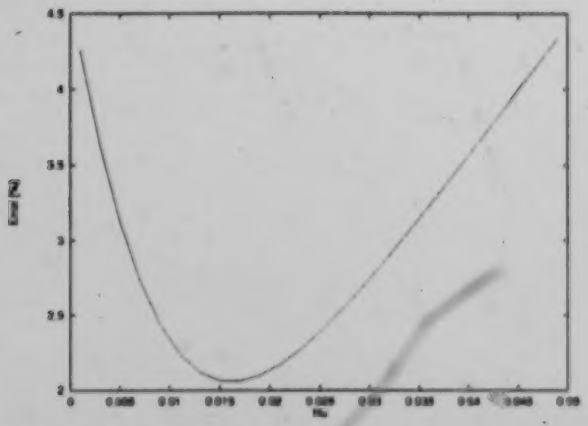


Figure 7.2: Optimal non-uniform discretization

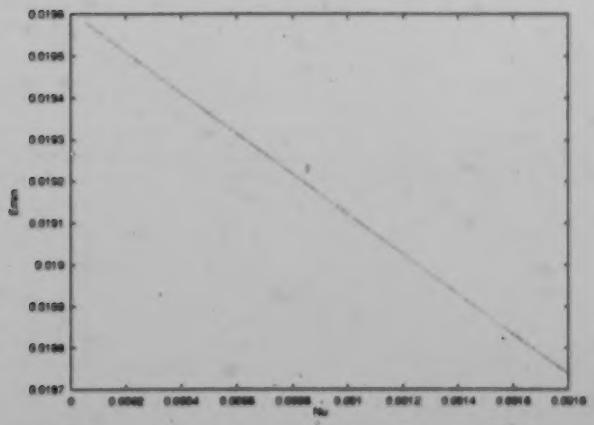


Figure 7.3: Changes in E_{min} with ν changes to keep constant overall resolution

Chapter 8

Stereo System with Cylindrical Sensors

In this chapter we study the depth estimation error in a stereo system using two cameras with cylindrical image planes. We demonstrate that such a stereo imaging system has superior performance in terms of the depth estimation error in comparison with the stereo imaging system using conventional cameras with flat image planes.

Section 8.1 describes the cameras that can be modeled as cameras with cylindrical image planes. In Section 8.2, we obtain the formula for calculating the depth of a 3D point using a stereo imaging system built with cylindrical sensors. Section 8.3 provides the error analysis in depth of a 3D point using cylindrical sensors. In Section 8.4, we compare the stereo imaging system built with cylindrical sensors against a system built with conventional cameras with flat image planes.

8.1 Cameras with Cylindrical Image Plane

Let us consider a device which consists of a single dimensional array of CCD sensors as opposed to conventional cameras which use CCD sensors that are two dimensional. In this device (illustrated in Figure 8.1), the CCD sensor is rotated around a center point. At each position the image projected onto the CCD sensor is a column of picture elements in the final image provided by the device. This device can be used to generate panoramic images of 360 degree rotation.

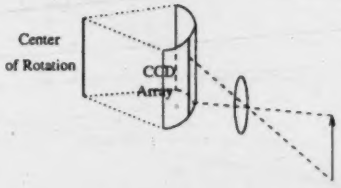


Figure 8.1: Model of the imaging system with single dimensional array

The camera described above is modeled to have an image plane of cylindrical form, and is therefore referred to as a *cylindrical camera*. The radius of the cylinder is the radius of rotation. The focal length of the lens in this device has significance only in the vertical direction. The form of the image in the horizontal direction is determined by the radius of rotation and the horizontal resolution.

8.2 Depth of a 3D Point

Figure 8.2 illustrates a stereo imaging system that uses the cylindrical cameras. In this stereo imaging system, we assume that the cameras have no independent tilt movement and are similar to an ordinary parallel stereo imaging system; the cameras are merely separated in the X direction. The focal length of each camera is denoted by f. The pixels on the sensor array are of a square form with size e. The pixel

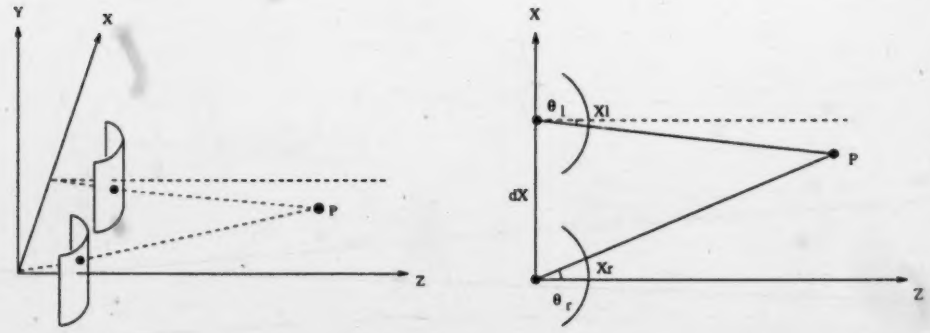


Figure 8.2: Stereo imaging system using cameras with cylindrical image plane

separation on the image plane in the horizontal and vertical directions are denoted by e_x and e_y, respectively. Note that e_y is always equal to e, but e_x is based on the steps and radius of the rotation. The radius of rotation (the distance between the focal point and the image plane) is denoted by R.

The separation distance between the cameras is denoted by dX. It is assumed that the origin of 3D world coordinates is at the focal point of the right camera. The projection of the 3D point $P(X, Y, Z)$ in the right and left cameras is represented by (x_r, y_r) and (x_l, y_l), respectively. Because of the discrete placement of pixels, these two projections are approximated to $(\hat{x}_r, \hat{y}_r)$ and $(\hat{x}_l, \hat{y}_l)$. The discretization error in turn leads to an estimate $(\hat{X}, \hat{Y}, \hat{Z})$ of the coordinates of point P.

The following theorem provides a formula for the calculation of depth (Z coordinate) of the point P in Figure 8.2.

Theorem 9 *The depth of a point P is calculated from the following formula:*

$$Z = \frac{dX}{\tan(x_r/R) - \tan(x_l/R)} \tag{8.1}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: In order to prove this theorem, let us consider the top view of our stereo

imaging system along the Y coordinate axis as illustrated in Figure 8.2(right). The distance between the projection of point P in the right camera's image plane to the center of the image along the X direction is an arc which is defined by the angle θ_r and represented by x_r. Similarly x_l is an arc defined by the angle θ_l.

Based on the arc length formula from elementary calculus we have:

$$x_r = R\arctan(X/Z) \tag{8.2}$$

$$x_l = R\arctan((X - dX)/Z) \tag{8.3}$$

Then we have:

$$Z = \frac{X}{\tan(x_r/R)} \tag{8.4}$$

$$Z = \frac{X - dX}{\tan(x_l/R)} \tag{8.5}$$

By eliminating X from the above equations we have:

$$Z = \frac{dX}{\tan(x_r/R) - \tan(x_l/R)}$$

—

8.3 Depth Estimation Error Analysis

In this section we analyze the error in the stereo system with cylindrical sensors. We obtain an upper bound on the maximum error in depth estimation for the 3D point P. The following theorem provides such an upper bound:

Theorem 10 *An upper bound on the relative error in depth of point $P(X, Y, Z)$ is obtained from the following formula:*

$$E_Z = \frac{e_x}{2R}\left(\frac{X}{Z} - \frac{X - dX}{Z}\right)$$
$$+Z\frac{e_x}{RdX}\left(1 + \frac{X(X - dX)}{Z^2}\right) \tag{8.6}$$

For definition of parameters refer to Section 3.1 and List of Symbols.

Proof: In order to prove this theorem, let us use the result of Theorem 1 in Section 3.1. The projection of point P in each camera has at most $e_x/2$ discretization error. We have:

$$\hat{x}_r = x_r \pm e_x/2 \qquad \hat{x}_l = x_l \pm e_x/2 \tag{8.7}$$

Using Theorem 1, the depth of point P is estimated as:

$$\hat{Z} = \frac{dX}{\tan(\hat{x}_r/R) - \tan(\hat{x}_l/R)} \tag{8.8}$$

The following trigonometric equation exists:

$$\tan(\alpha + \beta) = \frac{\tan(\alpha) + \tan(\beta)}{1 - \tan(\alpha)\tan(\beta)} \tag{8.9}$$

By substituting for $\hat{x}_r$ and $\hat{x}_l$ and using the above trigonometric equations, We have:

$$\hat{Z} = \frac{dX(1 \mp e_x/2R\tan(x_r/R))(1 \mp e_x/2R\tan(x_l/R))}{(\tan(x_r/R) \pm e_x/2R)(1 \mp e_x/2R\tan(x_l)/R) - (\tan(x_l/R) \pm e_x/2R)(1 \mp e_x/2R\tan(x_r)/}$$

By multiplying the parentheses together and ignoring the second order terms, we have:

$$\hat{Z} = \frac{dX(1 \mp e_x/2R(\tan(x_r/R) + \tan(x_l/R))}{(\tan(x_r/R) - \tan(x_l/R)) \pm (e_x/R + e_x/R\tan(x_r/R)\tan(x_l/R))}$$

Then we have:

$$\hat{Z} = Z\frac{(1 \mp e_x/2R(\tan(x_r/R) + \tan(x_l/R))}{1 \pm \frac{Ze_x}{RdX}(1 + \tan(x_r/R)\tan(x_l/R))} \tag{8.10}$$

For practical situations such as a room environment where Z is in the same order of magnitude as dX, the above equation can be rewritten as follows:

$$\hat{Z} = Z((1 \mp \frac{e_x}{2R}(\tan(x_r/R) + \tan(x_l/R)))(1 \mp \frac{Ze_x}{RdX}(1 + \tan(x_r/R)\tan(x_l/R)) \tag{8.11}$$

By multiplying two parentheses together and ignoring the second order terms we have:

$$E_Z = | \frac{\hat{Z} - Z}{Z} | \leq \frac{e_x}{2R}(\tan(x_r/R) + \tan(x_l/R)) + \frac{Ze_x}{RdX}(1 + \tan(x_r/R)\tan(x_l/R))$$

By replacing in the above equation for x_r and x_l from equations 8.2 and 8.3, equation 8.6 is obtained.

—

8.4 Comparison of Flat vs. Cylindrical Stereo

For the stereo imaging system using the cylindrical image plane, pan vergence movements have no effect on the stereo system configuration due to the nature of the cameras. The advantage of this new stereo imaging system is that it eliminates the negative effects of vergence movements on the depth estimation error.

Figure 8.3 illustrates the variation of depth uncertainty in a typical stereo imaging system when using cameras with flat image planes. Figure 8.4 illustrates the depth estimation error in a stereo imaging system built with cameras with cylindrical image planes. In both these figures, the pixel separations on both the flat and cylindrical image planes is the same. For the sake of simplicity, the radius of the cylindrical camera is chosen to be equal to the focal length of flat image plane cameras. As the figures illustrate, the error in a stereo imaging system with flat image plane and no vergence angle is roughly the same as the stereo imaging system with a cylindrical image plane. However, the error in the flat case grows larger with the vergence movements toward the object in the scene, while the error in the cylindrical case remains unchanged with vergence movements.

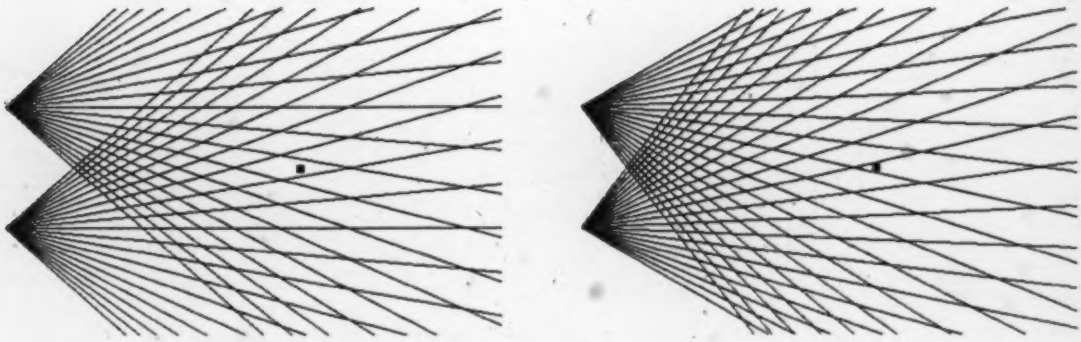


Figure 8.3: Variation of depth estimation error in stereo with vergence movements using cameras with flat image plane

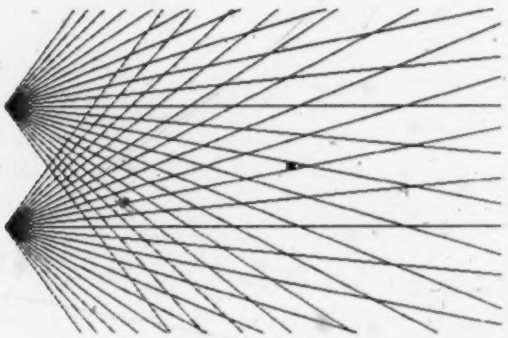


Figure 8.4: Depth estimation error in stereo using cameras with cylindrical image plane

While the focal angle for all the pixels of the cylindrical cameras is the same, the focal angles of the conventional cameras changes with the position of the pixels. Comparison of flat vs. cylindrical stereo is only fair when we consider the size of the focal angles for both flat and cylindrical case. The size of the focal angle in the cylindrical cameras is a function of the pixel size as well as the radius of the cylinder. Therefore, for any given conventional camera with a flat image plane, a cylindrical camera is comparable when the pixel sizes are the same and the radius is chosen such that the focal angles are equal to the average of the focal angles of the conventional flat image plane cameras.

Let us consider an stereo imaging system with a flat image plane and uniform resolution where f = 10 mm, 40 mm × 40 mm imaging area, e_x = 0.05 mm and dX = 100 mm. The object of interest is in the region $280 < Z < 320$ and $35 < X < 65$. Figure 8.5 illustrates the upper bound of depth estimation error as a function of vergence angle changes in the left and right cameras.

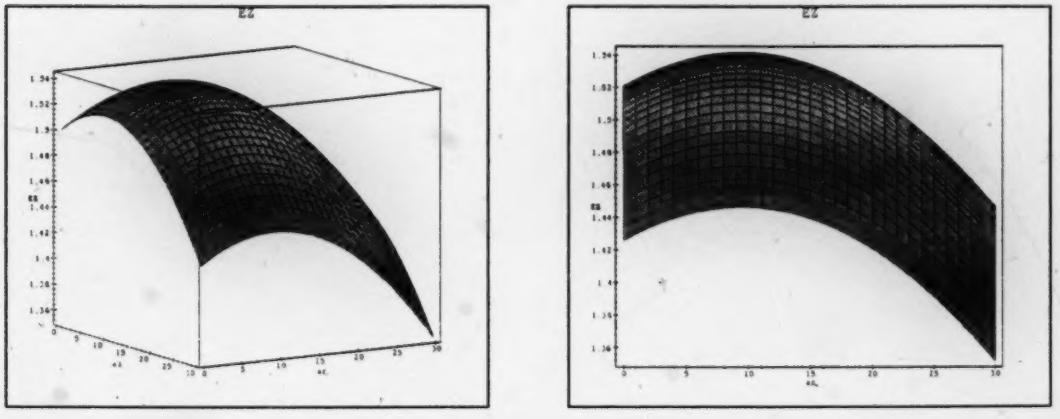


Figure 8.5: Upper bound on the depth estimation error versus the vergence angle in a stereo system with flat image plane and uniform resolution

Figure 8.6 illustrates the upper bound changes for the above stereo imaging system when the cameras have non-uniform exponential resolution described by the following equation:

$$e_x(i) = E_{min}e^{\gamma i} \tag{8.12}$$

and $E_{min} = 0.025$, $\gamma = 0.01$.

Figure 8.7 illustrates the depth estimation error in the above stereo imaging system when the cameras have cylindrical image planes. In order to have a fair comparison with flat image planes, the radius R is selected based on the size of the flat image planes as well as the focal length (in our example $R = 15.0$). As this figure

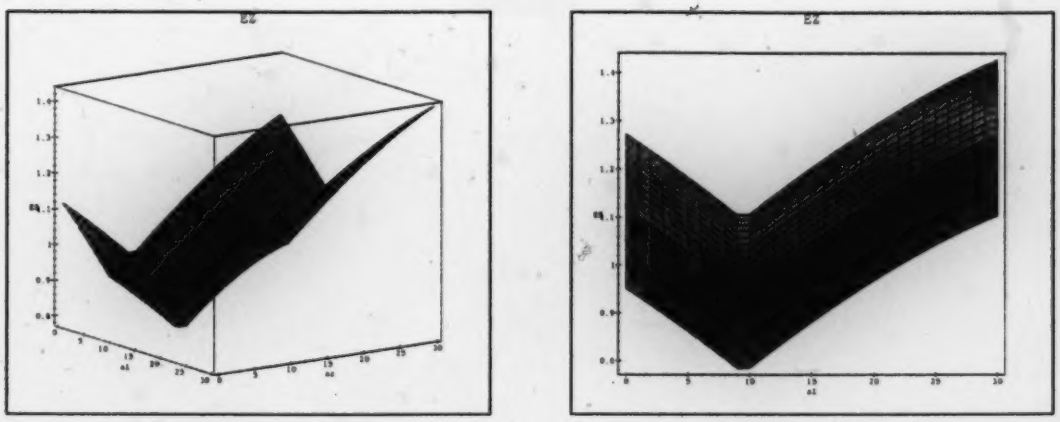


Figure 8.6: Upper bound of depth estimation error versus vergence angle in stereo system with flat image plane and non-uniform resolution

shows, the upper bound of the error is constant 1.02% and it is much more desirable than the other two cases.

Besides the fact that the upper bound of error in the cylindrical stereo is uniform with vergence movements, the cost of providing super high resolution using the rotating linear CCD cameras is less than conventional flat CCD cameras. As well, lens distortion (which could be a source of error in conventional cameras) is not a factor in rotating linear CCD cameras. In the above example, when the e_x is reduced to half (0.025) in cylindrical sensors, the upper bound of error which is proportional to e_x (based on equation 8.6), is reduced from 1.02% (illustrated in Figure 8.7) to 0.51%. Achieving such low errors with conventional cameras is costlier.

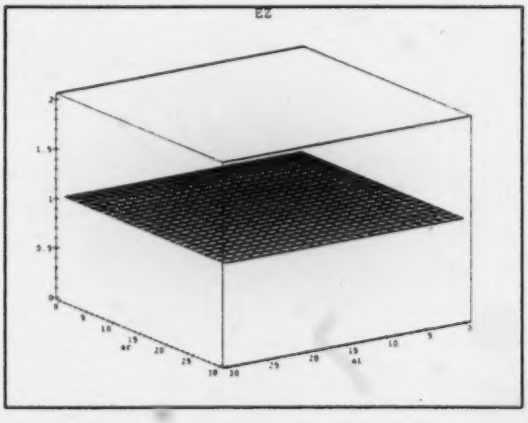


Figure 8.7: Upper bound of depth estimation error versus vergence angle in stereo system with cylindrical cameras

Chapter 9

Experimental Results

This chapter presents the results of experiments conducted to verify the theoretical results obtained in the previous chapters. In Section 9.1 a simulated stereo imaging system was used. In this system, average error for an object in the scene was studied. In Section 9.2 a real stereo imaging system using conventional cameras was used and the depth of selected 3D points in the scene was studied for different stereo imaging configurations. Two sets of tests were carried out in this experiment. In the first test, coplanar points in the scene were studied. In the second test, selected features in a 3D object (a Mr. Potato Head toy) were examined. In order to carry out these tests, a number of software tools were developed. In Section 9.3, a stereo imaging system with cylindrical cameras was used. The depth of selected points in the scene were studied.

9.1 Simulated Stereo Imaging System

Let us consider a simulated stereo imaging system with two cameras providing images from a scene. For the sake of simplicity, we assume that the cameras do not have

tilt movements and vergence pan movements are symmetrical for both cameras. For a virtual object in the scene, a grid of points is selected and the projections of each point in the camera's image plane are obtained. Based on the characteristics of the image plane, the projection is rounded to the nearest pixel. The depth of the point is calculated from the rounded position; the obtained depth is compared with the actual depth and percentage error is calculated. The errors for all the points are averaged to obtain the average error percentage for the object. In this experiment we are only concerned with discretization error and other sources of errors such as lens distortion are ignored.

Figure 9.1 illustrates the average depth estimation error for an object located within $25 < X < 75$ and $350 < Z < 400$ boundaries. The average error is obtained for all the points within the boundary with 1 mm distance. The stereo imaging system in this example consists of two uniform resolution cameras with a pixel separation of 0.075 mm, where the baseline distance is dX=100 mm, and the focal length of f=10 mm. As well the cameras have symmetrical vergence movements with angle α.

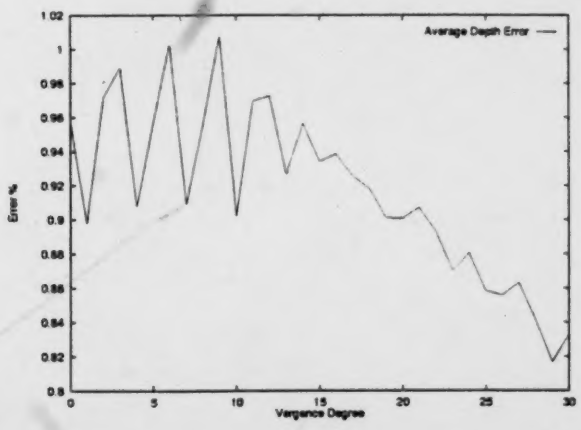


Figure 9.1: Average depth estimation error in a stereo system with uniform discretization

Figure 9.2 illustrates the average error for the same object of the above system and the stereo imaging system with the exact same configuration except that the cameras have non-uniform discretization of an exponential form $E_{min} = 0.05, \gamma = 0.01$.

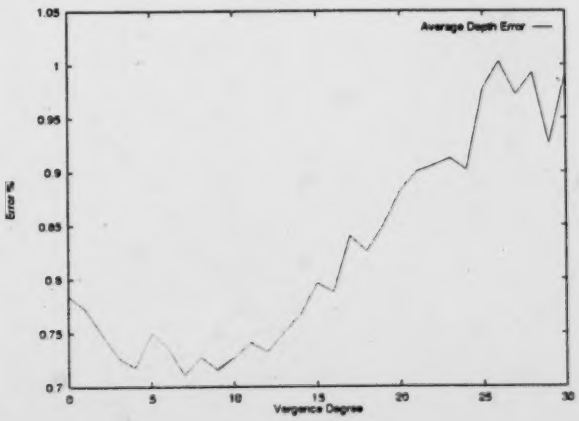


Figure 9.2: Average depth estimation error in a stereo system with non-uniform discretization.

Both figures clearly show the trend of changes in the average depth estimation error obtained by changes in the upper bound on the error. The irregularity in each case is due to the fact we are obtaining the real error and not an upper bound, and for each vergence angle there is the potential of having no error at all for some points; the projections of some points lie exactly on the pixel centers.

A similar experiment was conducted when cylindrical cameras were used instead of conventional flat image plane cameras. The R parameter was chosen to be 10 mm to be equal to the focal length of the previous experiment. As well, the pixel separation e was chosen to be 0.05 mm to provide a focal angle roughly equal to the average focal angle of the pixels in the flat uniform resolution case. Simulation indicates that the average error for the same object is 0.66; this is better than the cases

of uniform and non-uniform resolutions examined earlier. As well, the error unlike in the previous cases, remains constant with pan vergence movements of the cameras

9.2 Conventional Stereo Imaging System

9.2.1 Calibration Process

The first step to determine an accurate depth measurement in stereo imaging is the calibration process. This consists of determining the internal camera geometric and optical characteristics as well as its position and orientation respect to the world coordinate system.

We used Tsai's algorithm for camera calibration [37]. Tsai's algorithm is based on his comprehensive camera model, consisting of 11 parameters: six extrinsic parameters ($R_x, R_y, R_z, T_x, T_y, T_z$), which are the position and orientation of the camera coordinates with respect to the world coordinate system, and five intrinsic parameters (f, C_x, C_y, s_x, k_1) that describe the camera's image formation system. More details about Tsai's camera calibration technique are provided in Appendix B.

We require Tsai's exterior orientation parameters that describe the position and orientation of the camera to obtain the parameters in our stereo imaging model to use in the mathematical formulas obtained in the earlier chapters for calculating depth information. The required parameters are dx, α, β, γ, and θ.

Figure 9.3 describes how the parameters are calculated from calibration results. In order to obtain the parameters in our stereo configuration, we need to define the 3D coordinate system in our model based on the calibration 3D coordinate system. In our model, the 3D coordinates origin is in the focal point of the right camera. The

X axis of our coordinate system is obtained from:

$$\vec{X_M} = \vec{T_L} - \vec{T_R} \tag{9.1}$$

where $\vec{T_R}$ and $\vec{T_L}$ are the translation vectors of the right and left cameras, respectively. The Y axis of our coordinate system is arbitrarily chosen from the following equation:

$$\vec{Y_M} = \vec{X_M} \times \vec{Z_C} \tag{9.2}$$

Similarly, the Z axis is obtained from the following equation:

$$\vec{Z_M} = \vec{X_M} \times \vec{Y_M} \tag{9.3}$$

Based on calibration data, each camera orientation's vector is defined as follows:

$$\begin{aligned} \vec{X_{Cam}} &= \cos(r_x)\sin(r_y) \\ \vec{Y_{Cam}} &= \sin(r_x) \\ \vec{Z_{Cam}} &= \cos(r_x)\cos(r_y) \end{aligned} \tag{9.4}$$

dX is obtained from the length of the $\vec{X_M}$ vector. For the right camera the tilt angle β is the angle between the cameras orientation vector $\vec{P_{Cam}}$ and the $X_M Z_M$ plane. Therefore the first projection of the orientation vector $\vec{P_{Cam}}$ in the $X_M Z_M$ plane is obtained. β is calculated from the angle of the camera's orientation vector and its projection. α is calculated from the angle between the $\vec{P_{Cam}}$ and $\vec{Z_M}$. A similar method is used to find θ and γ for the left camera. The equations are:

$$\vec{P_{Cam}} = \vec{Cam} - \frac{\vec{Cam} \cdot \vec{Y_M}}{\vec{Y_M}^2} \vec{Y_M} \tag{9.5}$$

$$\beta = \arccos(\bar{P_{Cam}} / \bar{Cam}) \tag{9.6}$$

$$\vec{P_{ZCam}} = \frac{P_{Cam} \cdot Z_M}{\vec{Z_M}^2} \vec{Z_M} \tag{9.7}$$

$$\alpha = \arccos(\bar{P_{ZCam}} / \bar{P_{Cam}}) \tag{9.8}$$

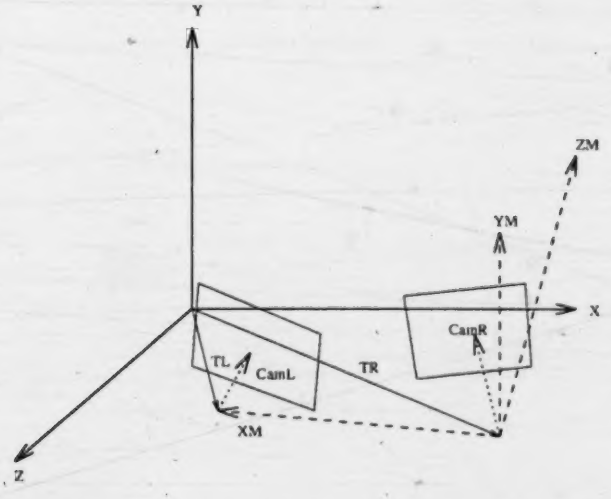


Figure 9.3: Calculation of our models parameters from calibration results

9.2.2 Coplanar Points

In this test, we use a set of coplanar points in the 3D scene and measure the depth of selected points using three different vergence positions.

Because we did not have access to cameras with non-uniform discretization, we manipulated the original images acquired from the cameras to mimic the non-uniform discretization. The depths of the same points were measured using the resulting images. The depth information obtained from non-uniform resulting images was

compared with the depth information obtained from uniform subsampled images to make the comparison a fair one. The subsampled images have almost the same number of pixels as the non-uniform resolution images.

Parameter	Value	Units
f	11.013987	mm
C_x	320.000000	pixels
C_y	240.000000	pixels
k_1	-2.137029e-03	$1/mm^2$
s_x	1.000000	
T_x	-61.721653	mm
T_y	-47.529553	mm
T_z	553.052224	mm
R_x	-22.488247	deg
R_y	-1.125861	deg
R_z	-3.532439	deg
Mean UIP	1.508903	pix

Table 9.1: Calibration parameters for one of the camera's positions

Point	*PositionA*	*PositionB*	*PositionC*
α	-1.32	6.94	14.49
β	-23.11	-22.75	-24.97
γ	-1.12	-6.91	-13.87
θ	-22.48	-22.52	-22.23

Table 9.2: Orientation information obtained from the calibration for the three different positions

For the calibration purposes, test points were created by applying Letteraset onto a flat block. A single point on the surface was selected as the origin of the 3D world coordinates. The camera was calibrated using a monoview set of 60 selected points which, because they lie within the block surface, will be coplanar. The coordinates of the selected data points were measured with submillimeter accuracy. Calibration was carried out by first acquiring a grey scale image of the scene; image was then thresholded to provide a binary image. In the resulting image, the positions

Figure 9.4: Original stereo images in position *A*

Figure 9.5: Original stereo images in position *B*

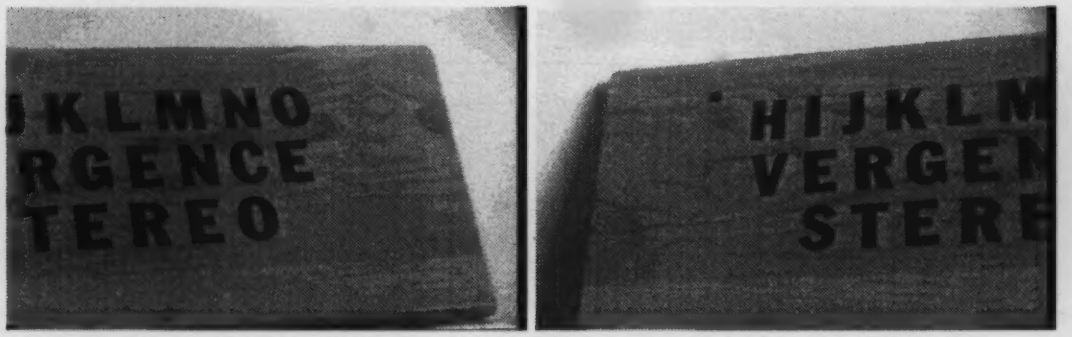


Figure 9.6: Original stereo images in position *C*

of data points were accurately measured. This information as well as the measured 3D coordinates of the data points, was passed to the calibration program.

The images were acquired using a Cannon CCD camera with a 1/4 inch image plane and 577 × 433 pixels on the CCD image plane. However, the frame grabber consists of 640 × 480 pixels. Stereo images were obtained by moving the camera to provide two shots of the same scene; the camera was calibrated for each view. Although the intrinsic parameters remain the same, the extrinsic parameters must be recomputed for each camera's position. The same procedure was repeated for each of the left and right images with three different camera positions (A, B, C), as illustrated in Figures 9.4, 9.5 and 9.6. Results of the calibration were then used to calculate the pan and tilt angles in our stereo model as depicted in Figure 6.1.

Table 9.1 illustrates a sample of the resulting calibration parameters obtained from one of the camera positions. The first 5 parameters in the table are the intrinsic ones and are considered to be independent of the position of the camera. The rest of the parameters are the extrinsic parameters and are dependent on the camera's position and orientation. The small value for mean Undistorted Image Plane (UIP) indicates that the calibrated model did a good job for capturing camera's behavior. Table 9.2 illustrates the various angle information obtained from calibration for the three different positions.

Point	Position A	Position B	Position C
1	597.07	596.27	585.11
2	599.24	600.65	582.11
3	574.82	568.59	593.07
4	552.97	552.32	542.09
5	543.25	538.74	540.34
6	541.40	542.12	538.43

Table 9.3: Depth of the selected points measured using the original images

Point	Position A	Position B	Position C
1	598.73 — 0.2 %	599.02 — 0.5 %	583.62 — 0.2 %
2	599.24 — 0 %	602.34 — 0.3 %	582.22 — 0.01 %
3	574.07 — 0.1 %	571.03 — 0.4 %	592.44 — 0.1 %
4	555.26 — 0.4 %	552.34 — 0.003 %	541.62 — 0.1 %
5	544.38 — 0.2 %	538.43 — 0.05 %	542.29 — 0.36 %
6	541.55 — 0.02 %	541.67 — 0.08 %	539.11 — 0.1 %

Table 9.4: Depth of the selected points and the percentage error measured using the subsampled images and compared with the results obtained from the original images

Point	Position A	Position B	Position C
1	595.29 — 0.3 %	596.12 — 0.02 %	588.87 — 0.64 %
2	595.96 — 0.5 %	600.20 — 0.07 %	585.19 — 0.5 %
3	571.72 — 0.5 %	568.93 — 0.05 %	594.08 — 0.17 %
4	548.76 — 0.8 %	552.23 — 0.01 %	545.43 — 0.6 %
5	541.12 — 0.4 %	539.44 — 0.1 %	543.93 — 0.66 %
6	538.53 — 0.5 %	541.69 — 0.08 %	542.04 — 0.67 %

Table 9.5: Depth of the selected points and the percentage error measured using the non-uniform images and compared with the results obtained from the original images

Table 9.3 illustrates the depth measured for six selected points in the scene using the original high resolution images with the three different vergence positions; the selected points were manually matched in both images. The selected points are the top left corners of the letters K and L in the first row, the two left corners of letter N in the second row, and the top left corners of letters E and R in the third row. Table 9.4 illustrates the depth calculated from the subsampled images and the percentage error when compared with the original resolution images. The original resolution images serve as the benchmark for accurate depth information. The error described in the table is over and above the error introduced due to the discretization in the original images. Table 9.5 illustrates the depth calculated from the non-uniform resolution images and the error in the depth. As was expected, the error for position B using non-uniform resolution images is considerably lower compared to the error

of depth using the uniform images; the error for positions A and C, however, are generally higher in the non-uniform resolution case.

9.2.3 Points on a 3D Object

In the previous test described in Section 9.2.2, we ignored the fact that even the original resolution images introduce error in the depth calculation due to their discretization effect. In this test we demonstrate this in a more tangible fashion to show how dimensions are estimated for a 3D object using stereo images.

In this test we use stereo imaging to measure depth of selected feature points on a Mr. Potato Head toy. We used the same camera and procedure for obtaining stereo images as in the last test, and once again used them in three different vergence positions. Depth of the feature points is studied on the subsampled images as well as the non-uniform resolution images.

For each set of images, edge detection was first performed on the images. Then a number of well-defined feature points in the images were selected from the resulting binary images. The features from both images were then manually matched. Epipolar computations (described in chapter 6) were used as a guide to make the matching process easier. Then the depth for each selected point was calculated.

The following figures illustrate the images in three different vergence positions. As well, the subsampled images and the non-uniform resolution images in vergence position B are illustrated. For a larger illustrations of the subsampled and non-uniform resolution images, refer to Appendix C.

The following figures illustrate the selected feature points in Mr. Potato Head's face for which depth is calculated, as well as selected dimensions of Mr. Potato Head's face.

Figure 9.7: Original stereo images position A

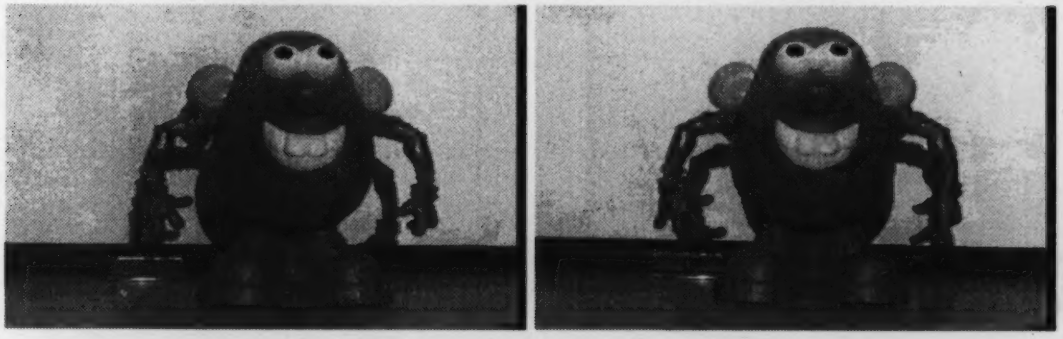


Figure 9.8: Original stereo images position B

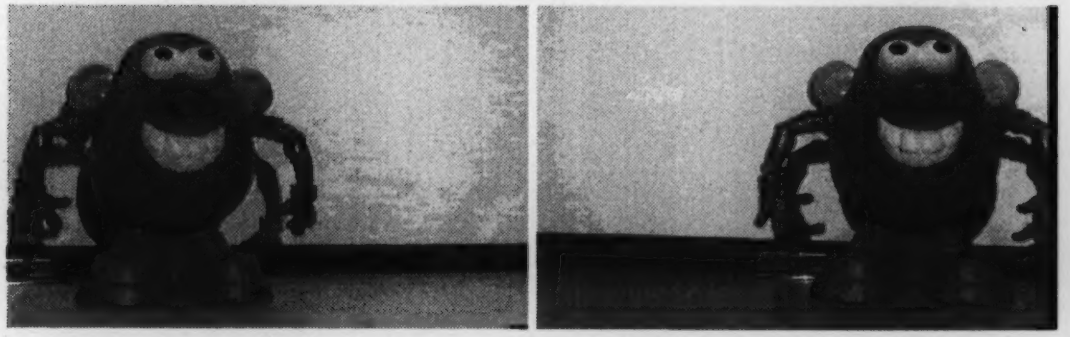


Figure 9.9: Original stereo images position C

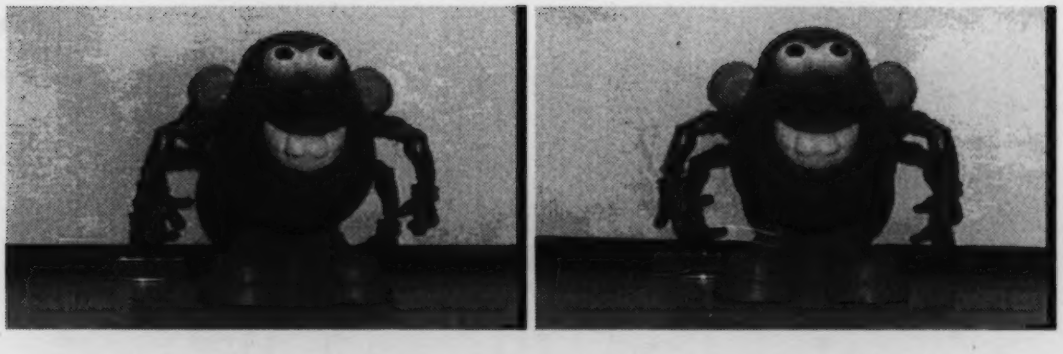


Figure 9.10: Original stereo images with uniform discretization, position B

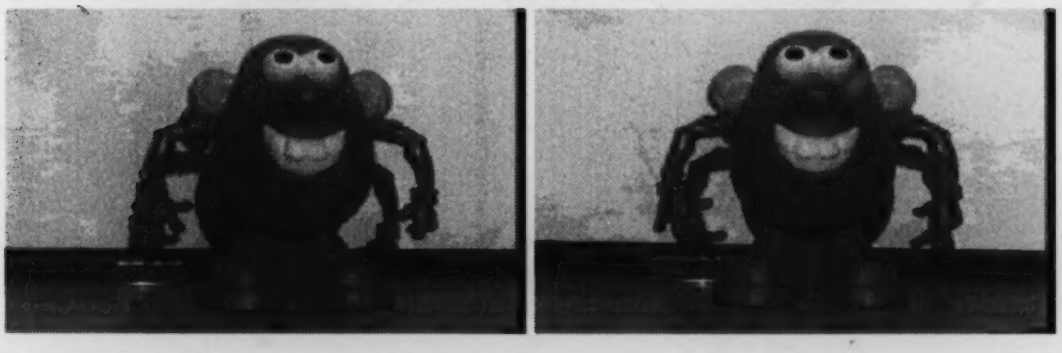


Figure 9.11: Subsampled stereo images with uniform discretization, position B

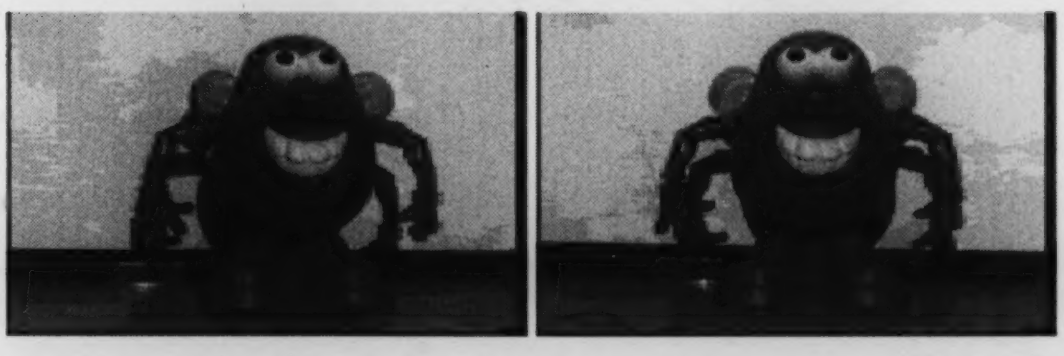


Figure 9.12: Stereo images with non-uniform exponential discretization, position B

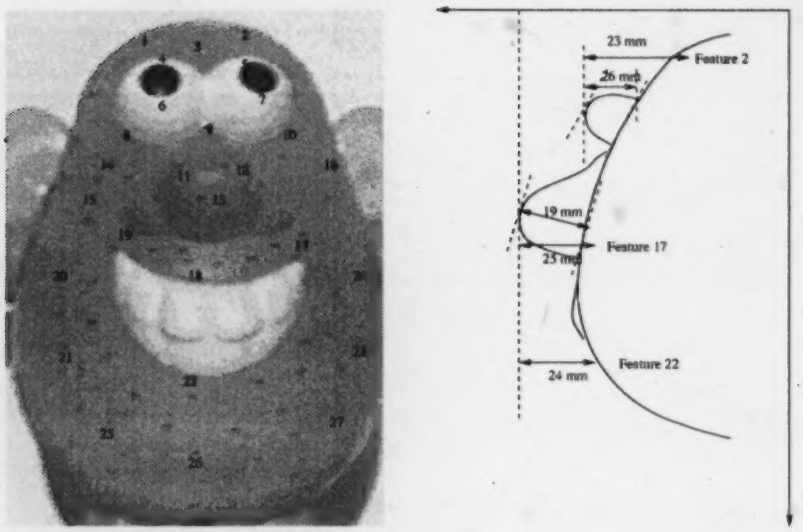


Figure 9.13: Mr. Potato Head's Dimensions (right), Mr. Potato Head's selected features (left)

For each set of images the depth of selected feature points were measured. Then the depth difference between some of the feature points is compared with the actual dimensions of Mr. Potato Head's face. This gives us a clear view of how accurate the depth estimations are. This information are recorded in a table for each set of images. Note that the errors are in estimation of dimensions, and that a small error in depth estimation can lead to high error in the dimension.

By examing the results it is obvious that, even in the original resolution images, there is error in estimating the dimensions of the 3D surface, and by considering the depth values, we can estimate the depth estimation error in the order of few percentage points.

In the subsampled images, the errors in the estimation of dimensions grow for almost all locations. In the non-uniform resolution case, the error in estimated dimensions depends on the location of the features and the position of images. In

Point	Position A	Position B	Position C
1	410.49	415.52	408.26
2	412.47	415.27	411.75
3	412.47	415.26	413.60
4	412.47	425.13	422.38
5	412.47	424.23	422.69
6	424.07	423.03	423.24
7	424.07	424.32	422.12
8	412.47	417.52	414.78
9	418.33	416.46	415.17
10	414.44	416.18	415.01
11	427.83	427.22	424.93
12	427.83	427.97	424.63
13	427.83	426.03	425.38
14	416.39	414.61	415.00
15	414.44	414.58	414.87
16	418.33	413.51	411.47
17	415.26	413.42	414.67
18	420.26	416.67	415.86
19	420.25	415.12	415.77
20	412.47	413.20	412.28
21	412.47	412.25	412.88
22	412.47	414.51	415.60
23	412.47	412.38	411.67
24	414.44	411.15	410.29
25	412.47	412.78	413.06
26	410.49	413.16	409.88
27	414.44	411.12	410.42

Table 9.6: Depth of the selected feature points using the original images

Features	Actual Distance	Error % Pos. A	Error % Pos. B	Error % Pos. C
Features 2,7	23	13.58	9.05	10.37
Features 17,13	25	12.57	12.61	10.71
Features 13,22	24	15.36	11.52	9.78

Table 9.7: Depth distance between the selected feature points using the original images, and percentage error compared to the actual dimensions

Point	Position A	Position B	Position C
1	404.48	410.95	412.32
2	410.49	418.36	414.86
3	404.48	414.24	420.81
4	410.49	419.99	429.11
5	410.49	420.91	423.52
6	422.17	418.98	428.02
7	422.17	426.17	426.41
8	404.48	411.22	420.82
9	416.39	418.37	415.12
10	410.49	418.12	420.35
11	422.17	423.17	428.69
12	422.17	429.84	432.41
13	427.83	424.58	431.19
14	410.49	415.75	416.98
15	410.49	417.65	421.69
16	416.39	414.14	417.63
17	416.39	413.53	422.13
18	416.39	416.94	420.44
19	416.39	408.11	419.68
20	410.49	415.15	412.29
21	408.22	410.21	409.48
22	414.48	420.76	417.80
23	410.49	412.06	403.61
24	410.49	408.97	414.59
25	410.49	416.96	418.59
26	404.48	410.91	418.39
27	410.49	403.24	418.39

Table 9.8: Depth of the selected feature points using the subsampled images

Features	Actual Distance	Error % Pos. A	Error % Pos. B	Error % Pos. C
Features 2,7	23	11.68	7.81	11.55
Features 17,13	25	11.44	11.05	9.06
Features 13,22	24	13.35	3.82	13.39

Table 9.9: Depth distance between the selected feature points using the subsampled images, and percentage error compared to the actual dimensions

Point	Position A	Position B	Position C
1	417.11	420.76	397.49
2	422.17	419.38	401.63
3	427.06	412.24	407.48
4	422.17	426.44	419.42
5	417.16	420.41	416.96
6	429.51	421.34	417.42
7	432.71	426.26	420.80
8	419.94	416.26	407.49
9	427.06	415.34	403.11
10	422.52	416.16	407.25
11	434.84	425.83	412.29
12	438.44	425.68	421.57
13	430.84	427.51	421.01
14	427.49	415.64	407.31
15	416.33	413.82	406.51
16	428.18	415.67	400.51
17	422.11	415.37	409.01
18	428.08	417.45	409.11
19	424.22	415.85	409.09
20	420.20	412.44	400.49
21	422.11	415.37	405.07
22	426.24	414.04	405.64
23	418.61	418.68	411.31
24	427.49	408.70	401.58
25	417.87	415.13	403.99
26	418.03	412.62	403.67
27	420.20	410.28	408.61

Table 9.10: Depth of the selected feature points using the non-uniform images

Features	Actual Distance	Error % Pos. A	Error % Pos. B	Error % Pos. C
Features 2,7	23	10.54	6.88	19.17
Features 17,13	25	8.73	12.14	12.0
Features 13,22	24	8.6	10.47	15.37

Table 9.11: Depth distance between the selected feature points using the non-uniform images, and percentage error compared to the actual dimensions

the case of position B and features in the foveal region such as 13 and 17, the result is good in comparison with the original images. However, for other positions, the results are severely distorted due to lower resolution outside of the foveal region.

9.3 Cylindrical Stereo System

In this section, we present the result of test performed with cylindrical cameras. A stereo system with the cylindrical image plane cameras was set up. Similar experiments as previous sections were conducted.

The cylindrical camera used in this experiment is (shown in Figure 9.14) manufactured by TelePhotogenics Inc. The resolution of the camera starts from around 4,000 pixels vertically to 40,000 pixels horizontally (for up to 360 degree field of view); models for higher resolution are available for other applications. For our experiments, we obtained the images by moving the same camera to two known and fixed positions in order to eliminate the possibility of variations between two cameras with similar specifications. In this case we have a lot more resolution in the images which we expect reduction in the depth estimation error.

Calibration of cylindrical cameras is beyond the scope of this thesis and we did not develop a calibration model. Instead of the calibration process, all the parameters of the camera relevant to depth estimation are accurately measured. Table 9.12 provides the parameters of this stereo imaging system which are required for depth calculation.

A few points in an indoor scene, as illustrated in Figure 9.15, were selected. The coordinates of the projections of each selected point in the scene, as well as the depth of each point were calculated from equation 8.1. Approximate physical

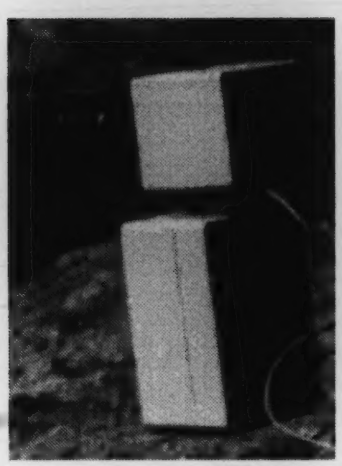


Figure 9.14: Camera used for cylindrical stereo imaging

Parameter	Value	Units
f	50.00	mm
e_x	14.9	micron
e_y	8	micron
R	39	mm
dX	150	mm

Table 9.12: Stereo system parameters for depth calculation

measurements have shown these estimates to be fairly accurate.

In order to further analyze the depth estimation in the cylindrical stereo system, we consider Mr. Potato Head in the scene. A number of features from Mr. Potato Head's face are selected; these are illustrated in Figure 9.17. Figure 9.18 illustrates the dimensions of the actual Mr. Potato Head, measured with submillimeter accuracy. Our focus of attention is on depth differences between various features.

Table 9.14 describes the coordinates of projections for each feature point as well as the result of depth calculation. Note that, the doll itself is not in a symmetrical position compared to the World Z axis, this results in slight depth differences between symmetrical features (e.g. feature 11 and 12).

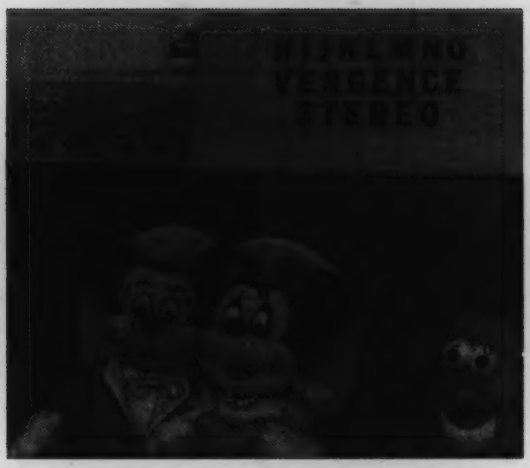


Figure 9.15: Points at which depth was estimated

Table 9.15 illustrates the depth differences between a number of selected features and their comparison with the actual dimensions. The result is quite good and the calculated dimensions are within 8% error in comparison with the actual dimensions. Note that the error percentage in the Table 9.15 is the error in dimensions; the error in depth for each feature is a lot less in term of percentage. With an accurate calibration model for the cylindrical cameras, higher precisions should be easily achievable.

Point	x_l,y_l	x_r,y_r	Depth
A	1679,182	1336,175	995.330597
B	2705,172	2357,175	992.890013
C	2285,518	1970,513	980.392840
D	1964,567	1647,568	972.379860
E	2529,567	2212,565	970.502177
F	2779,2098	2726,2046	630.481629
G	2967,2135	2913,2071	619.172822
H	2859,2333	2824,2265	611.081426
I	2853,2419	2833,2348	610.734019
J	2708,2494	2646,2434	645.218059
K	3047,2524	2987,2446	625.903207
L	1457,1765	1321,1785	761.553990
M	1615,1787	1490,1800	765.884373
N	occluded	1278,1951	N/A
O	1578,1960	1469,1976	738.364787
P	1435,2293	1340,2320	701.218996
Q	793,1697	651,1739	639.386019
R	1124,1753	977,1782	637.757062
S	836,1842	706,1893	635.843647
T	1009,1879	882,1920	638.653054
U	883,2182	746,2237	644.650064

Table 9.13: Depth of the selected points measured using the stereo images

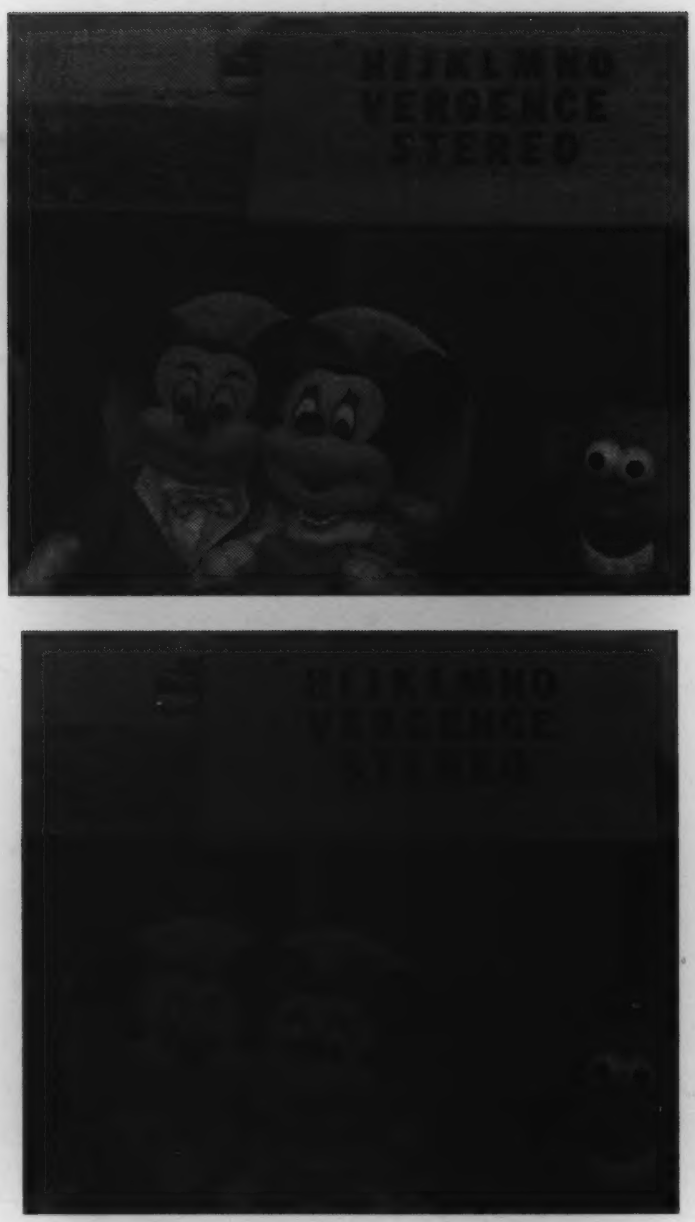


Figure 9.16: Stereo images, right image (top) and left image (bottom), using the cameras with cylindrical image planes

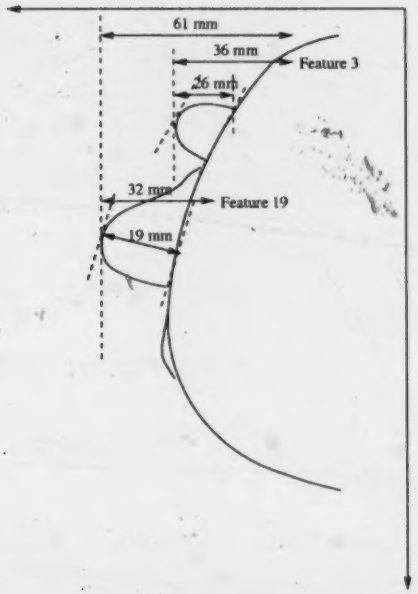


Figure 9.17: Mr. Potato Head's Dimensions

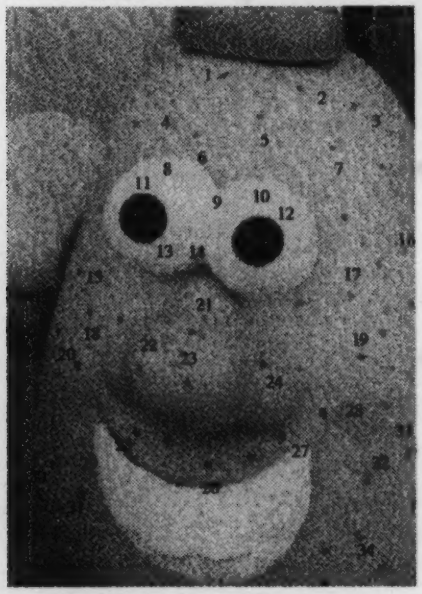


Figure 9.18: Mr. Potato Head's selected features.

Point	x_l,y_l	x_r,y_r	Depth
1	2908,1898	2812,1850	677.36
2	3036,1918	2924,1862	676.91
3	3124,1950	3000,1890	676.58
4	2820,1938	2736,1894	674.86
5	2972,1906	2872,1906	671.84
6	2868,1998	2780,1950	672.90
7	3096,2018	2980,1958	670.72
8	2828,2030	2748,1982	668.34
9	2900,2086	2812,2034	667.75
10	2988,2066	2890,2014	666.18
11	2776,2098	2724,2050	640.36
12	2968,2134	2894,2966	638.00
13	2776,2174	2724,2122	640.37
14	2880,2218	2816,2154	639.90
15	2668,2118	2604,2078	668.93
16	3172,2142	3048,2070	666.20
17	3128,2214	3012,2142	664.07
18	2690,2298	2616,2246	679.22
19	3148,2370	3040,2294	648.58
20	2670,2386	2598,2338	679.01
21	2862,2330	2816,2262	620.98
22	2780,2394	2746,2326	618.78
23	2856,2414	2816,2354	614.94
24	2978,2386	2916,2314	621.42
25	2768,2502	2708,2442	651.24
26	2884,2554	2828,2482	629.51
27	3008,2510	2944,2430	618.63
28	3068,2466	3000,2386	612.58
29	2680,2522	2604,2462	683.08
30	2628,2550	2560,2502	678.62
31	2680,2602	2608,2554	677.81
32	2640,2626	2580,2570	667.08
33	3138,2582	3060,2494	637.06
34	3090,2702	3000,2610	636.13

Table 9.14: Depth of the selected feature points on Mr. Potato Head measured using the stereo images.

Features	Depth Diff	Actual Dist.	Error %
12-23	638.00-614.94=23.06	25	8%
3-12	676.58-638.00=38.58	36	7 %
19-23	648.58-614.94=33.64	32	5 %
10-12	666.18-638.00=28.18	26	8 %

Table 9.15: Depth difference of the selected feature points on Mr. Potato Head measured using the stereo images.

Chapter 10

Conclusions and Future Work

In this thesis we studied depth estimation error in stereo imaging. First we studied a conventional stereo imaging system, and we established an upper bound on the depth estimation error. The behavior of this upper bound was studied with vergence movements of the stereo cameras. This study demonstrated that for objects in the scene depending on their location and the optical characteristics of the camera vergence may increase the depth estimation error. Then non-uniform discretization was proposed for stereo imaging. Non-uniform discretization in the horizontal direction as well as in both the horizontal and vertical directions was studied. Non-uniform discretization shows promising behavior in depth estimation error with vergence movements of the cameras. However, at least in an exponential model, there is no value of γ that led to uniform focal angles across the image plane. The benefit of uniform focal angles is that they would lead to constant upper bounds on the error with vergence movements. Cylindrical sensors were introduced; they have uniform focal angles across the image plane and therefore the upper bound on depth estimation error remains constant with vergence movements. This is a remarkable result and it eliminates the

negative effect of vergence movements on depth estimation error.

The analysis in this thesis could be used in developing optimal stereo imaging systems based on the results obtained. The cylindrical sensors can be used to build the system where negative effects of vergence are eliminated. As well non-uniform scheme can be combined with cylindrical approach to reduce the required processing power on the image.

Further study is required in the area of non-uniform resolution to study new schemes that may possibly lead to uniform focal angles on the flat image planes. As well non-uniform resolution sensors based on the schemes described in this thesis need to be developed. Developing the non-uniform sensors could be accomplished by changing the structure of sensor arrays in CCD cameras or it can be accomplished by designing optical devices to map the scene projection on the uniform sensor array in a non-uniform fashion.

Future work is also required to develop a comprehensive calibration scheme for cylindrical sensors to provide the intrinsic and extrinsic parameters for the system. There are potentially other type of image planes such as elliptical cameras that could lead to interesting behavior in stereo systems and deserves future attention.

References

[1] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *International Journal of Computer Vision*, pages 107–131, 1987.

[2] H. S. Baird. *Model-Based Image Matching Using Location.* MIT Press, 1985.

[3] D. A. Ballard and C. M. Brown. *Computer Vision.* Prentice Hall, 1982.

[4] A. Basu. Optimal discretization for stereo reconstruction. *Pattern Recognition Letters*, 13(11):813–820, November 1992.

[5] A. Basu and X. Li. A framework for variable resolution vision. In *Advances in Computing Information - ICCI 91*, pages 721–732, Ottawa, Canada, May 1991.

[6] A. Basu and S. Licardie. Modeling fish-eye lenses. In *IEEE IROS Conference*, Yokohama, Japan, July 1993.

[7] A. Basu and H. Sahabi. Analysis of depth perception with vergence and spatially varying sensing. In *ECCV Workshop on Natural and Artificial Visual Sensors*, Stockholm, Sweden, May 1994.

[8] A. Basu and H. Sahabi. Optimal non-uniform discretization for stereo recon-

struction. In *IEEE International Conference on Pattern Recognition*, Vienna, Austria, August 1996.

[9] A. Basu and H. Sahabi. Analysis of depth estimation error for cylindrical stereo imaging. In *Submitted for IEEE Intl Conf on Robotics and Automation*, San Fransisco, U.S.A, April 2000.

[10] S. D. Blostein and T. S. Huang. Error analysis in stereo determination of 3d point positions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(6):752–765, November 1989.

[11] R. A. Brooks. Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence*, 17:285–348, 1981.

[12] Andrea Califano, Rick Kjeldsen, and Ruud M. Bolle. Data and model driven foveation. In *Proceedings of IEEE International Conference on Pattern Recognition*, pages 1–7, June 1990.

[13] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.

[14] J. Clark and N. Ferrier. Modal control of visual attention. In *Proceedings of the International Conference on Computer Vision*, pages 521–531, Tarpon Springs, Florida, December 1988.

[15] L.S. Davis. A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4(3):248–270, 1976.

[16] U. R. Dhond and J. K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, November 1989.

[17] W. E. L. Grimson. A computer implementation of a theory of human stereo vision. *Phil. Transactions of the Royal Society of London*, B292:217–253, 1981.

[18] W. E. L. Grimson and E. C. Hildreth. Comments on digital step edges from zero-crossings of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):121–126, January 1985.

[19] M. J. Hannah. Bootstrap stereo. In *Proceedings of ARPA Image Understanding Workshop*, pages 201–208, College Park, MD, April 1980.

[20] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):121–136, February 1989.

[21] Y. C. Kim and J. K. Aggarwal. Positioning 3-d objects using stereo images. *IEEE Journal of Robotics and Automation*, RA-3(4):361–373, August 1987.

[22] D. Marr. *Vision*. W. H. Freeman Company, 1982.

[23] D. Marr and E. Hildreth. Theory of edge detection. *Proc. R. Soc. Lond.*, pages 187–217, 1980.

[24] D. Marr, G. Palm, and T. Poggio. Analysis of a cooperative stereo algorithm. *Biological Cybernetics*, 28:223–229, 1978.

[25] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of Royal Society of London*, B204:301–328, 1979.

[26] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Navigation*, RA-3(3):239–248, June 1987.

[27] J. E. W. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

[28] E. S. McVey and J. W. Lee. Some accuracy and resolution aspects of computer vision distance measurements. *IEEE Trans on Pattern Analysis and Machine Intelligence*, PAMI-4(6):646,649, November 1982.

[29] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics and Image Processing*, 31:2–18, 1985.

[30] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of 5th International Joint Conference of Artificial Intelligence*, page 584, 1977.

[31] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):139–154, March 1985.

[32] T. J. Olson and D. J. Coombs. Real-time vergence control for binocular robots. *International Journal of Computer Vision*, pages 881–888, 1991.

[33] RCA. Rca sid51232 technical note. Technical report, RCA, 1975.

[34] J. J. Rodriguez and J. K. Aggarwal. Quantization error in stereo imaging. In *Proc. of IEEE CVPR*, pages 153–158, Ann Arbor, Michigan, June 1988.

[35] H. Sahabi and A. Basu. Analysis of error in depth perception with vergence and spatially varying sensing. *Computer Vision and Image Understanding*, 63(3):447–461, May 1996.

[36] G. Sandini and V. Tagliasco. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14:365–372, 1980.

[37] R.Y. Tsai. A versatile camera caliberation technique for high accuracy 3d machine vision metrology. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.

[38] J. Weng, N. Ahuja, and T. S. Huang. Matching two perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):806–825, August 1992.

[39] Z. Yang and Y. F. Wang. Error analaysis 0f 3d shape construction from structured lighting. *Pattern Recognition*, 29(2):189–206, February 1996.

Appendix A

Imaging Device

Modern cameras are based on solid state sensor arrays composed of discrete silicon imaging elements, called photosites, that have voltage output proportional to the intensity of the incident light. The technology used in solid state sensors is based on charge-coupled devices (CCD), where each photosite is a MOS capacitor structure.

In order to understand the CCD technology, we examine the architecture of a commercial CCD device, RCA SID51232 [33]. This device generates a standard interlaced 525-line television picture. In this device, the cell size is 0.03 mm × 0.03 mm, and the image area size is 7.31×9.75. The useful spectral response range extends from 420 nm to 1100 nm.

Figure A.1 illustrates the block diagram of the SID51232, which consists of the following subsections:

- Image Area: The image area is an array of analog CCD shift registers containing 320 parallel columns of 256 sensing cells. Each cell is a MOS N-channel device defined by three adjacent polysilicon gate electrodes in the vertical direction and adjacent channel stops in the horizontal direction. The gates in each cell are

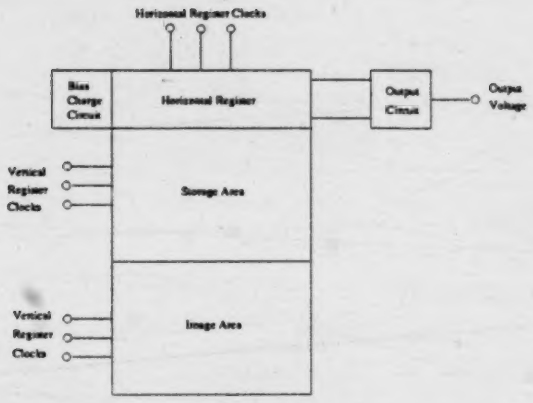


Figure A.1: Block diagram of a CCD imaging device.

connected in parallel with the corresponding gates in other cells. These three connections are called vertical register clocks $\Phi_{VA1}, \Phi_{VA2}, \Phi_{VA3}$. The transfer electrodes are made of polysilicon, which is transparent to most wavelengths and therefore there is no opaque part in the image area to cause picture details to be obscured.

When a pulse waveform is applied to the vertical register clocks, a light image focused on this register is integrated into a charge pattern of electrons during high part of the pulse and transferred to the storage area during the vertical blanking period.

- Storage Area: The storage area has the same architecture as the image area and contains the same sized array of cells, the cells hold the previous image frame to allow conversion of a charge pattern image into a sequential horizontal readout. The storage area is shielded from external light.

  The storage area is clocked with the image area during vertical blanking to transfer the complete image from image area to the storage area. During the

horizontal blanking, the entire charge pattern will be propagated one cell toward the horizontal register.

- Horizontal Register: The horizontal register is an output register and it has the same three phase structure as the image and storage areas. The horizontal register receives one line of picture information from the storage area during the horizontal blanking interval. The register contains the 320 active elements corresponding to the 320 columns in the image and storage areas.

- Bias Charge Circuit: The bias charge circuit provides a uniform low-noise bias charge to the output of the horizontal register to maximize the horizontal resolution.

- Output Circuit: The output circuit extracts the CCD signal from the horizontal register.

Appendix B

Tsai's Calibration Algorithm

Tsai's camera model is based on 11 parameters: six extrinsic parameters ($R_X, R_Y, R_Z, T_X, T_Y, T_Z$) and five intrinsic parameters (f, C_X, C_Y, s_X, k_1). The five intrinsic parameters for each camera are constant; the extrinsic parameters need to be recomputed once the camera moves.

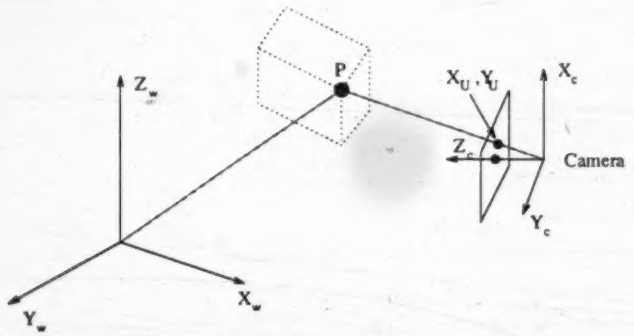


Figure B.1: Tsai's camera model

In Tsai's model illustrated in Figure B.1, the X_C, Y_C, Z_C axes describe the camera's coordinate system, in which the (X_C, Y_C) plane is parallel to the camera's image plane and the Z_C axis coincides with the camera's optical axis.

The following equation describes the relationship between the world coordinate

system (X_M, Y_M, Z_M) and the camera's coordinate system (X_C, Y_C, Z_C):

$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = R \begin{pmatrix} X_W \\ Y_W \\ Z_W \end{pmatrix} + T \tag{B.1}$$

where R denotes the rotation transformation which can be described as three sequential rotation transformations around X, Y, and Z axes by angles R_X, R_Y, and R_Z. Also, T describes the translation transformation.

The following equation describes the relationship between the coordinates of a projection point (X_U, Y_U) in the camera's image plane, with its coordinates in the camera's coordinate system (X_C, Y_C, Z_C) and the focal length f of the camera:

$$X_U = f\frac{X_C}{Z_C} \tag{B.2}$$

$$Y_U = f\frac{Y_C}{Z_C} \tag{B.3}$$

Due to lens distortion there is a displacement from the actual position of the projection (X_d, Y_d) to its ideal position (X_U, Y_U). K_1 is the parameter which describes the radial lens distortion. We have:

$$X_U = X_d(1 + K_1\rho^2) \tag{B.4}$$

$$Y_U = Y_d(1 + K_1\rho^2) \tag{B.5}$$

$$\rho = sqrt((X_d)^2 + (Y_d)^2)) \tag{B.6}$$

The final element of Tsai's model is the difference between the position of the projection point on the image plane (X_d, Y_d) and its coordinates in the camera's frame buffer (X_f, Y_f).

$$X_f = (e_x)^{(} - 1)X_d s_x + C_x \tag{B.7}$$

$$Y_f = (e_y)^{(} - 1)Y_d + C_y \tag{B.8}$$

where C_X and C_Y are the coordinates of the intersection of the Z_C axis and the camera's image plane, e_x and e_y are the pixel separations in X and Y directions, respectively, and s_x is a scaling factor denoting the ratio between the number of elements in the image plane and the number of pixels in the camera's frame buffer in the X direction.

Appendix C

Stereo Images

In Chapter 9, Figures 9.11 and 9.12 illustrate the subsampled and non-uniform resolution images of Mr. Potato Head. Due to the size and quality of those images, the differences might be hard to notice. Therefore in this appendix we include larger images to highlight the differences. The following figures illustrate the left image in position B for the original, subsampled, as well as the non-uniform cases.

Figure C.1: Left original resolution image

Figure C.2: Left subsampled resolution image

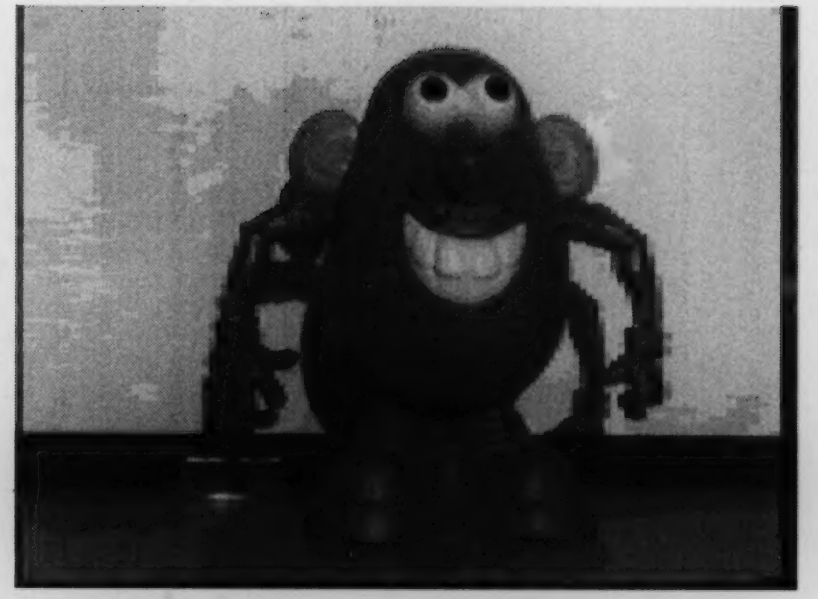


Figure C.3: Left non-uniform resolution image